

Institute of Computing for Climate Science

# To Bayesian Optimisation and Beyond Gaussian Processes as Decision Makers

Henry Moss







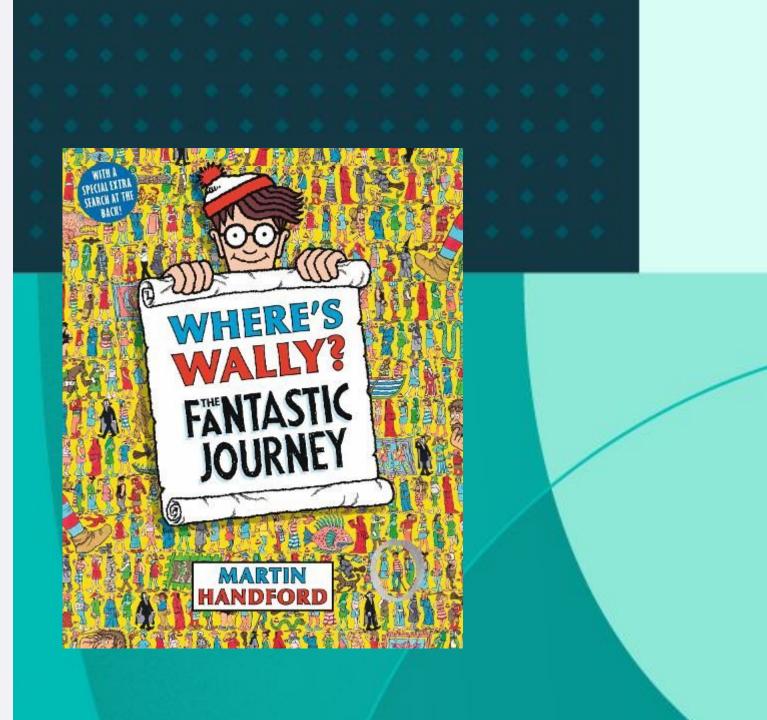
# **Bayesian Search**







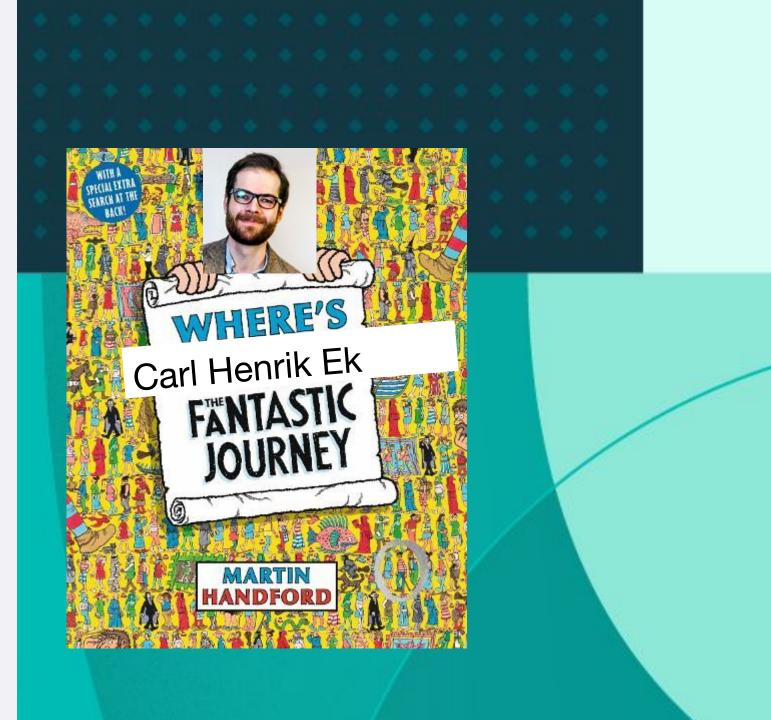
# **Bayesian Search**





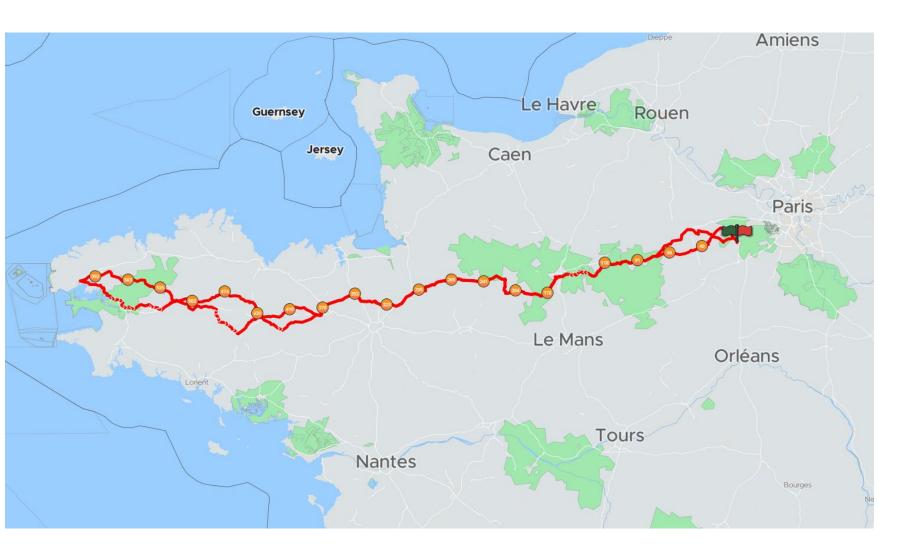


## **Bayesian Search**



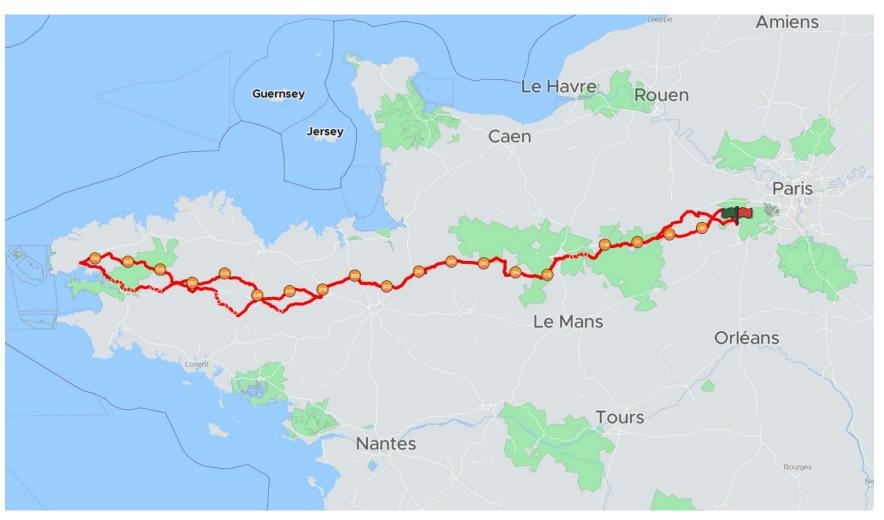


٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠
٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠
٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠

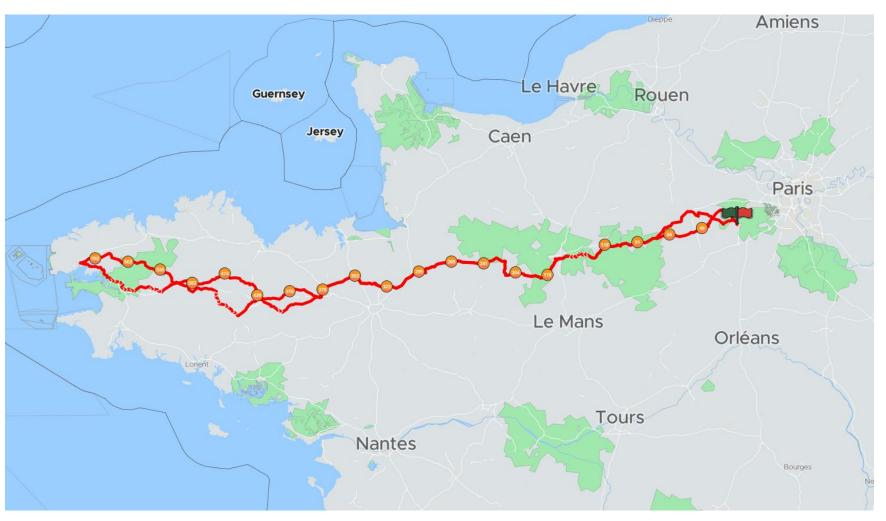




٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠
٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠
٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠

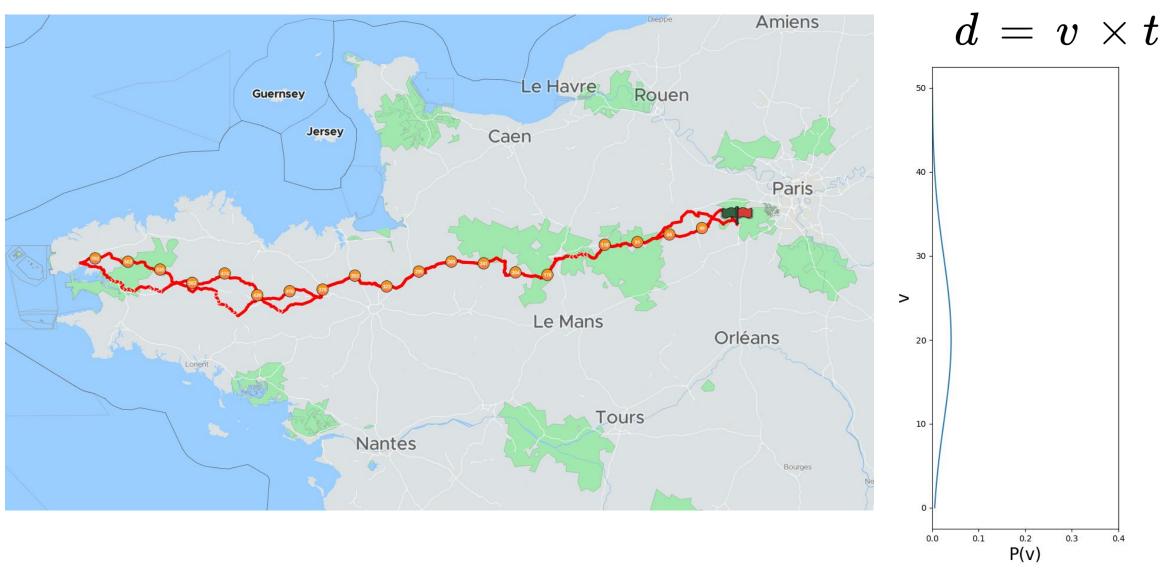




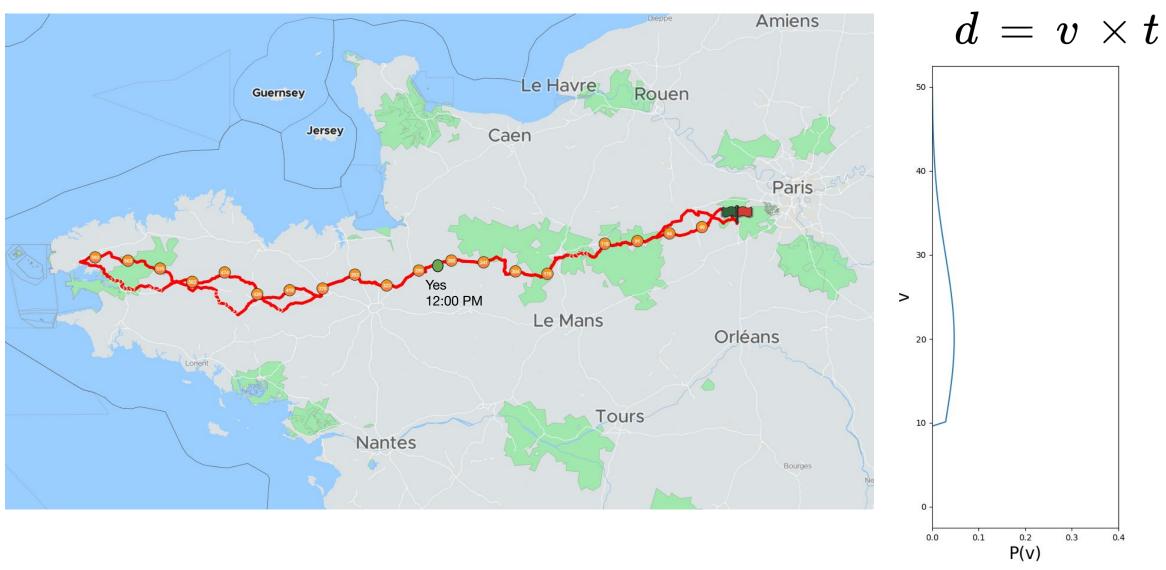


 $d\,=\,v\, imes t$ 

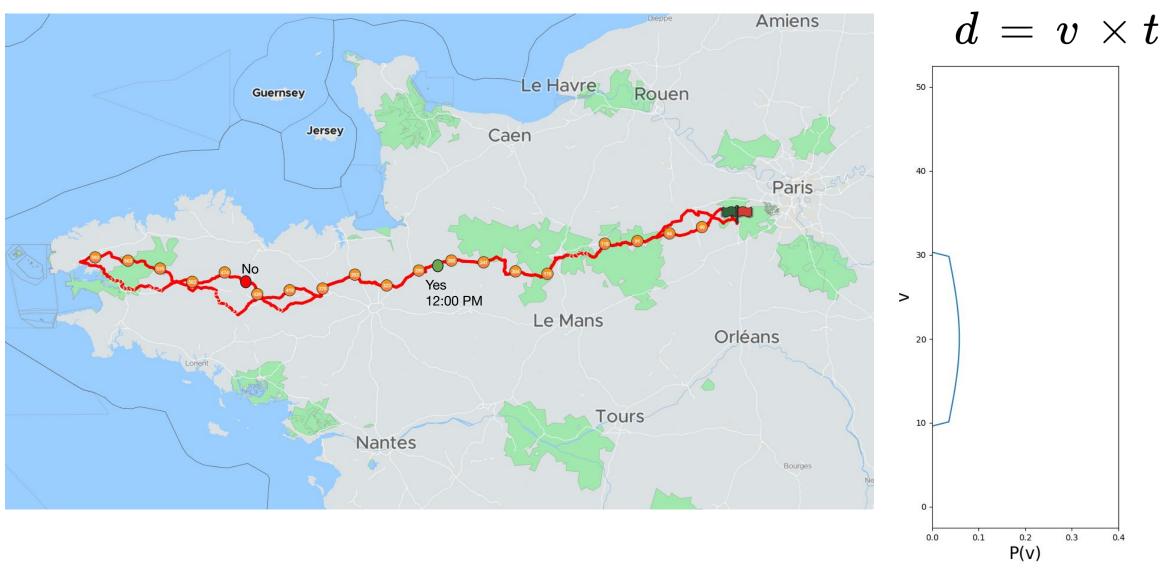




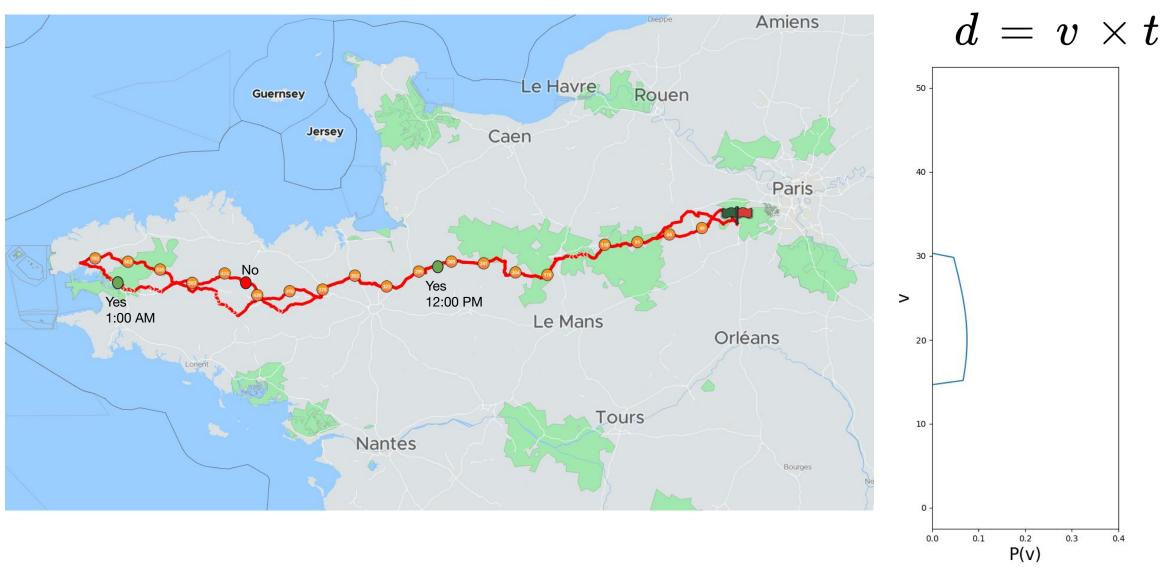




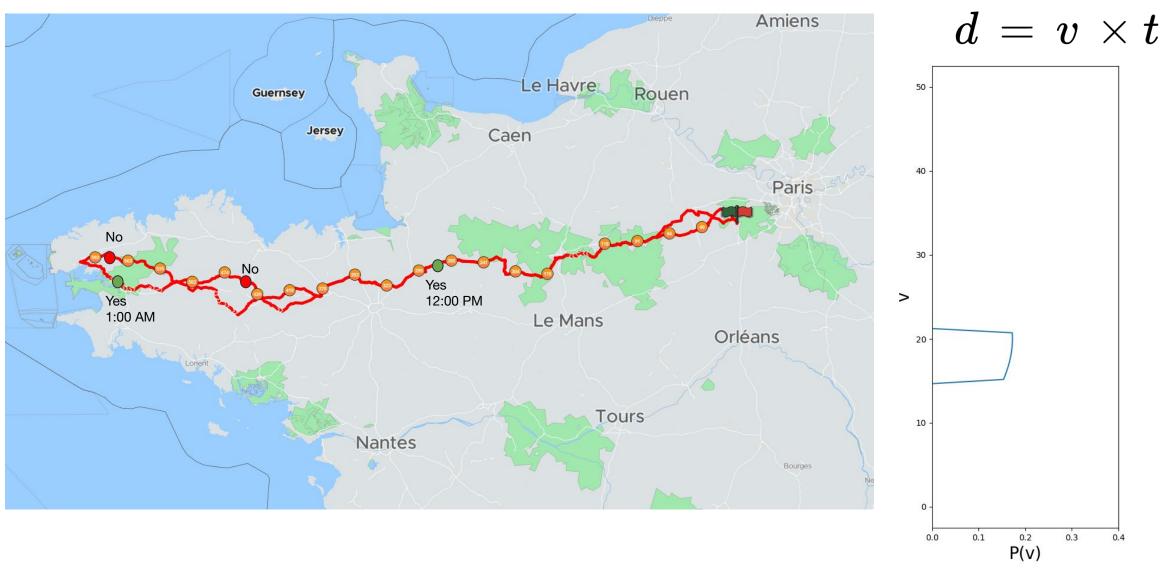










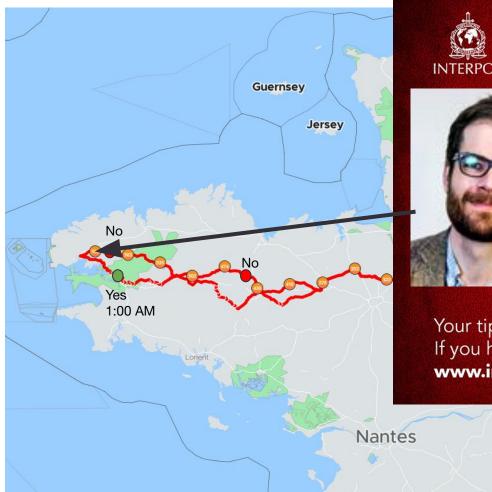






### Where is Carl Henrik?

At 3:30 AM?





Your tip could be the missing piece in the puzzle. If you have any information, contact your local police or go to www.interpol.int > Wanted persons

Bourges

Tours



At 3:30 AM?



But can we do better than **random**???



Institute of Computing for Climate Science

# What is Active Learning?

Bayesian search for learning functions









Let's make use of uncertainty estimates to make better models

Collect initial data

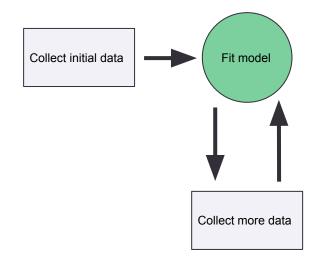






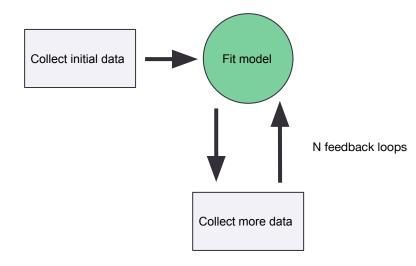






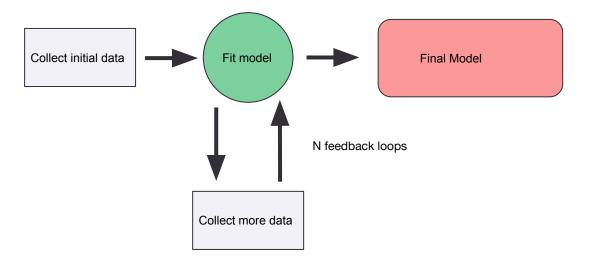




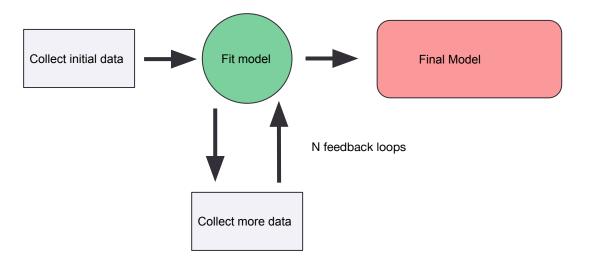


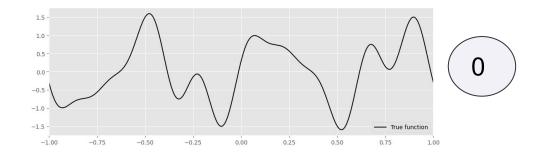




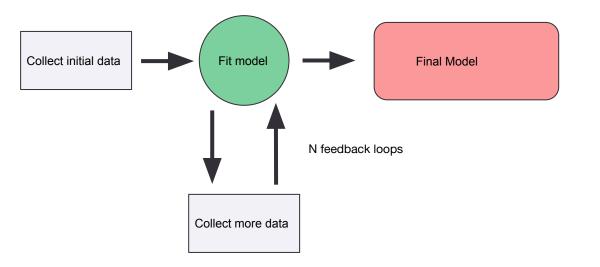


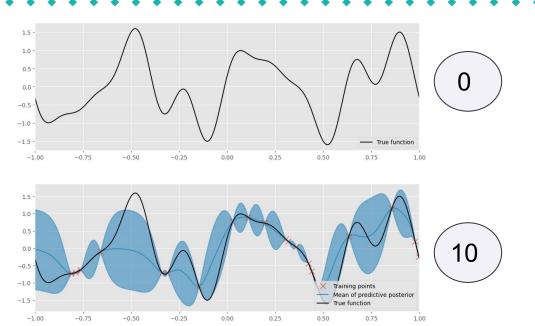
### Sequential data collection



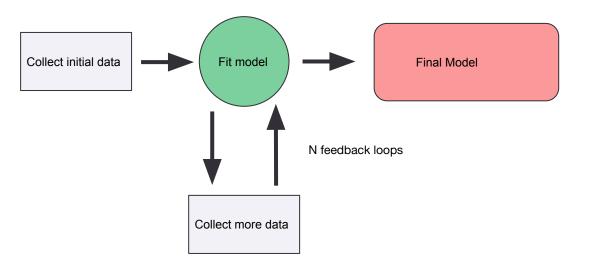


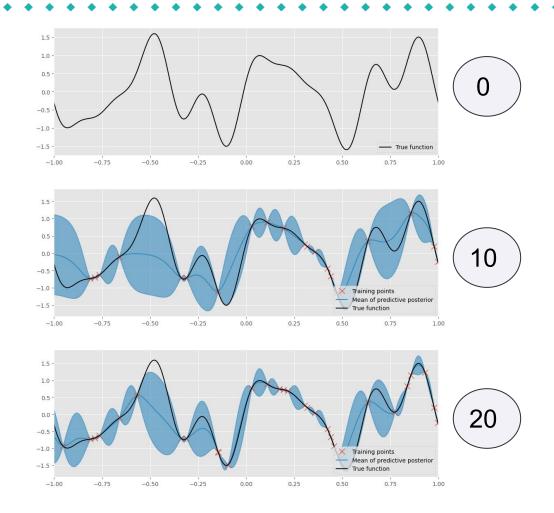
### Sequential data collection



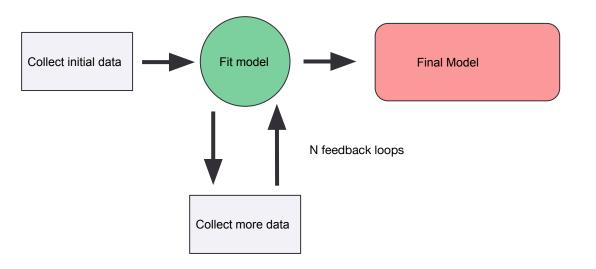


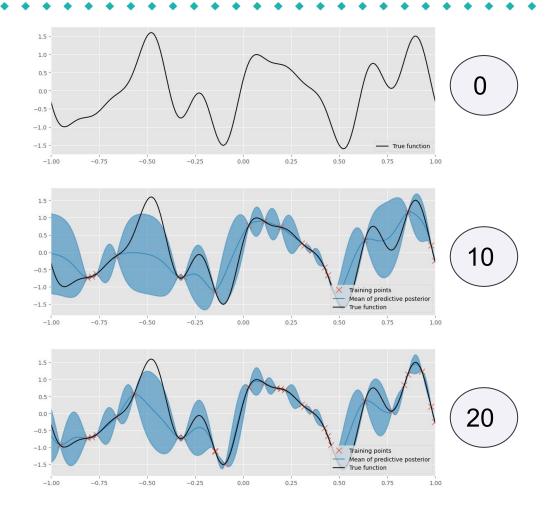
#### Sequential data collection

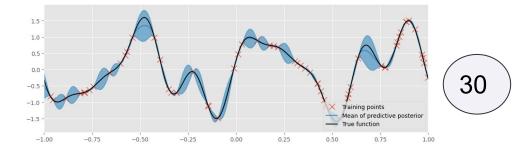




#### Sequential data collection

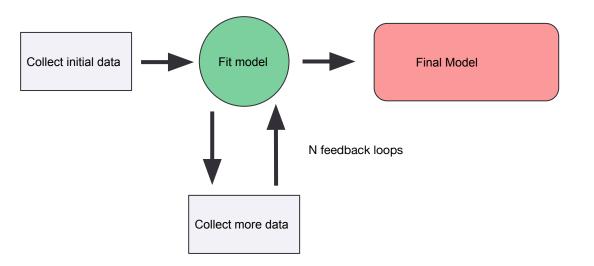


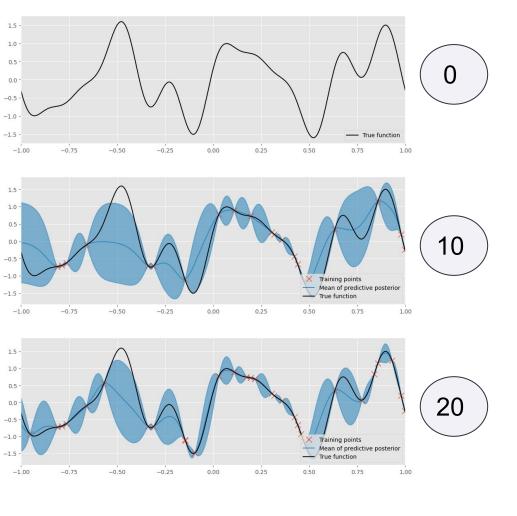




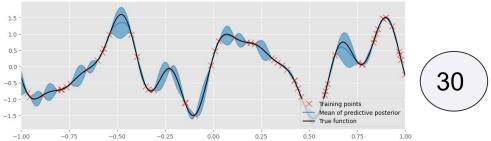
#### Sequential data collection

Let's make use of uncertainty estimates to make better models





#### But can we do better than **random**???







Sequentially collecting more data to improve your model for the task at hand

• I care about **regression** —> collect data to improve global model accuracy



- I care about **regression** —> collect data to improve global model accuracy
- I care about the **maximum** value of my process —> collect data in promising regions (Bayesian Optimisation)



- I care about **regression** —> collect data to improve global model accuracy
- I care about the **maximum** value of my process —> collect data in promising regions (Bayesian Optimisation)
- I'm interested in **multiple objectives** -> populate the Pareto front (Multi-objective Bayesian Optimisation)

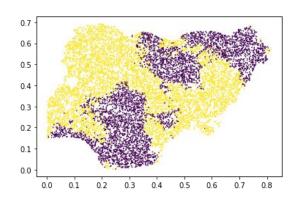


- I care about **regression** —> collect data to improve global model accuracy
- I care about the **maximum** value of my process —> collect data in promising regions (Bayesian Optimisation)
- I'm interested in **multiple objectives** -> populate the Pareto front (Multi-objective Bayesian Optimisation)
- I care about predicting a threshold -> choose data close to threshold (level-set design)

### Active learning

Sequentially collecting more data to improve your model for the task at hand

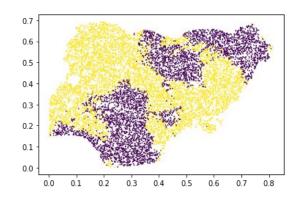
- I care about **regression** —> collect data to improve global model accuracy
- I care about the **maximum** value of my process —> collect data in promising regions (Bayesian Optimisation)
- I'm interested in **multiple objectives** -> populate the Pareto front (Multi-objective Bayesian Optimisation)
- I care about predicting a **threshold** -> choose data close to threshold (level-set design)



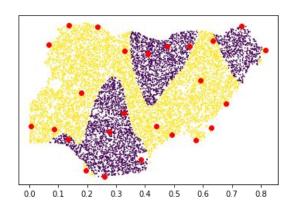
Malaria incidence in Nigeria

### Active learning

- I care about **regression** —> collect data to improve global model accuracy
- I care about the **maximum** value of my process —> collect data in promising regions (Bayesian Optimisation)
- I'm interested in **multiple objectives** -> populate the Pareto front (Multi-objective Bayesian Optimisation)
- I care about predicting a threshold -> choose data close to threshold (level-set design)



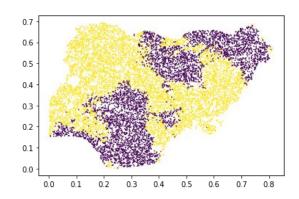
Malaria incidence in Nigeria



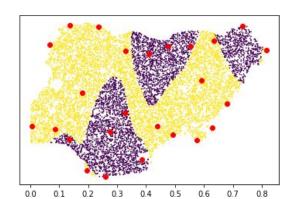
Model on Random data

### **Active learning**

- I care about **regression** —> collect data to improve global model accuracy
- I care about the **maximum** value of my process —> collect data in promising regions (Bayesian Optimisation)
- I'm interested in multiple objectives -> populate the Pareto front (Multi-objective Bayesian Optimisation)
- I care about predicting a **threshold** -> choose data close to threshold (level-set design)



Malaria incidence in Nigeria



0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8

Model on Random data

Model from data chosen by Active learning





# So, Bayesian Optimisation?

i.e. Active learning for optimisation





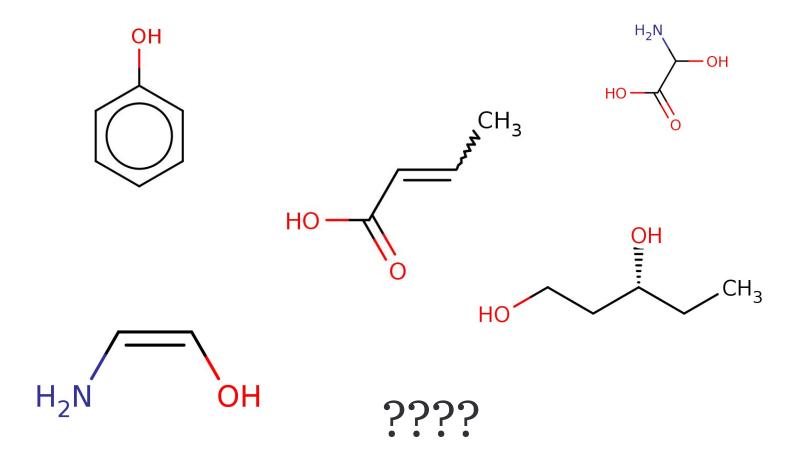
٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	•
٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	•
٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	•

# A molecular design pipeline

# A molecular design pipeline

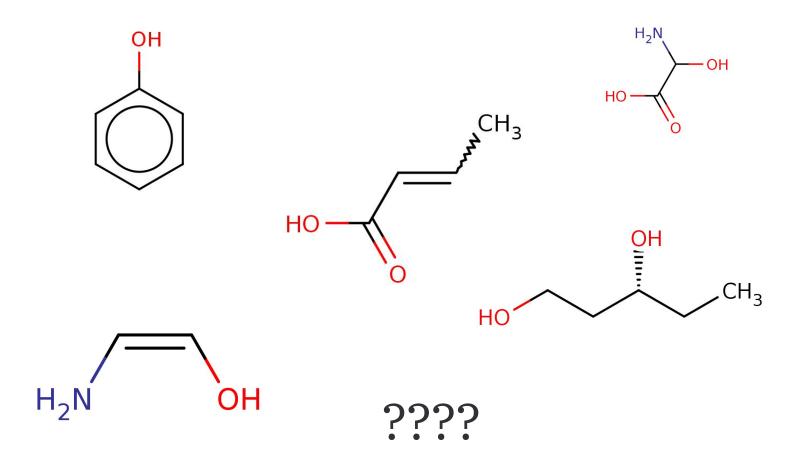
Efficiently explore molecule space

• Large library of candidates



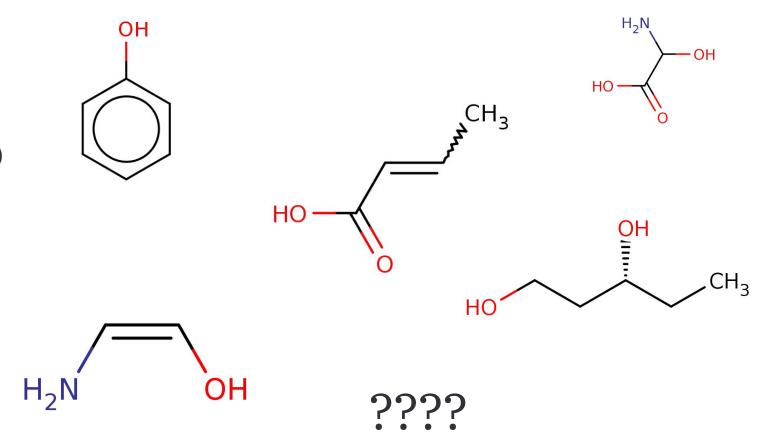
# A molecular design pipeline

- Large library of candidates
- **Expensive** experiments (<10)



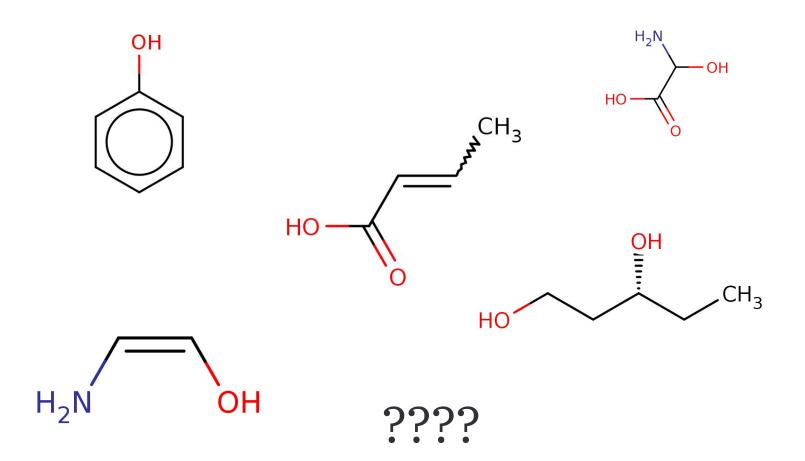
# A molecular design pipeline

- Large library of candidates
- Expensive experiments (<10) (IN A LAB !!!)



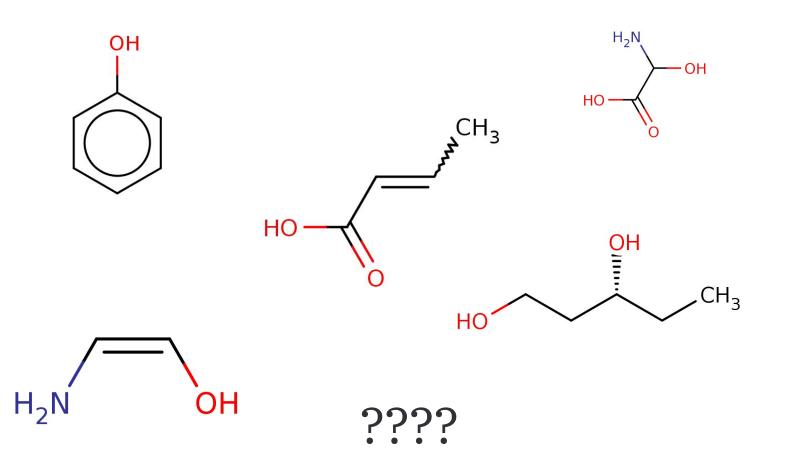
# A molecular design pipeline

- Large library of candidates
- **Expensive** experiments (<10)
- High degree of **parallelism**



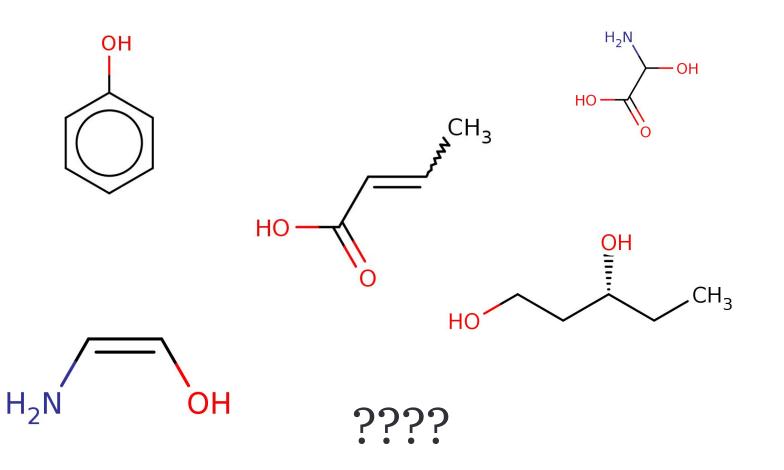
# A molecular design pipeline

- Large library of candidates
- **Expensive** experiments (<10)
- High degree of parallelism
- Want molecules with high affinity



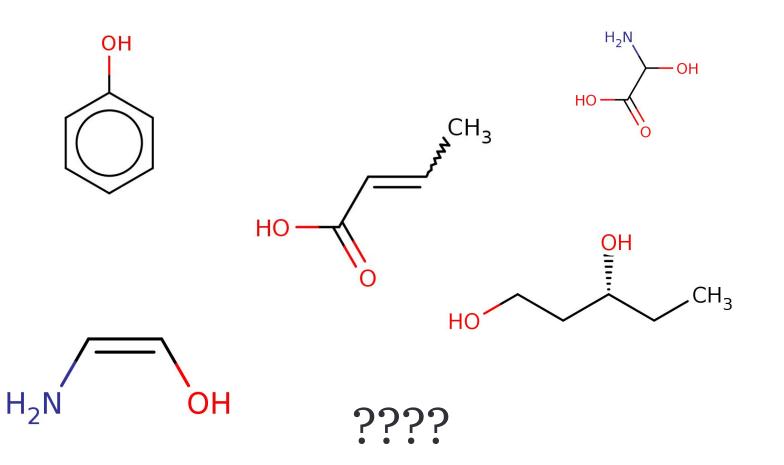
# A molecular design pipeline

- Large library of candidates
- **Expensive** experiments (<10)
- High degree of parallelism
- Want molecules with high affinity
  - Also easy to make



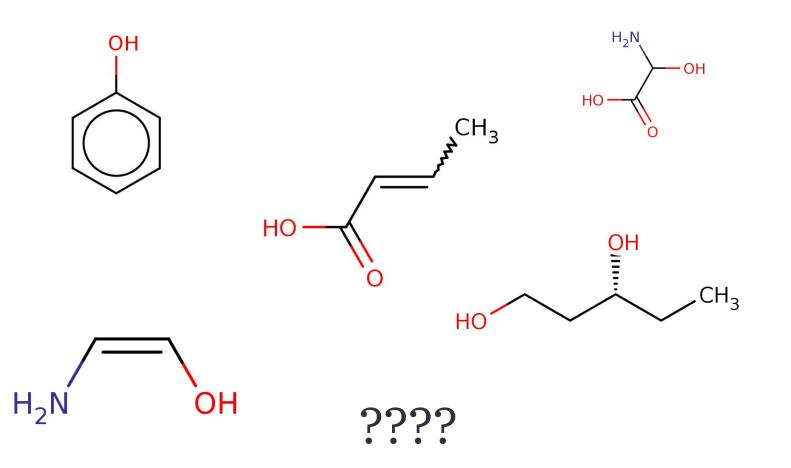
# A molecular design pipeline

- Large library of candidates
- **Expensive** experiments (<10)
- High degree of **parallelism**
- Want molecules with high affinity
  - Also easy to make
  - $\circ~$  Don't stick to themselves



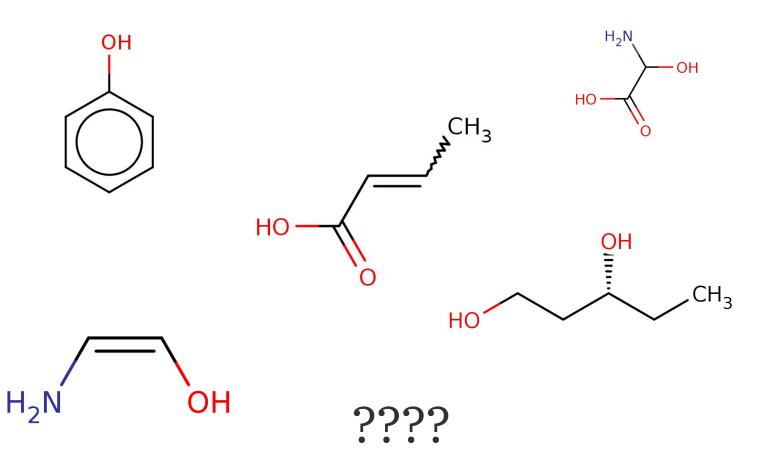
# A molecular design pipeline

- Large library of candidates
- **Expensive** experiments (<10)
- High degree of **parallelism**
- Want molecules with high affinity
  - Also easy to make
  - Don't stick to themselves
  - $\circ$  Stable



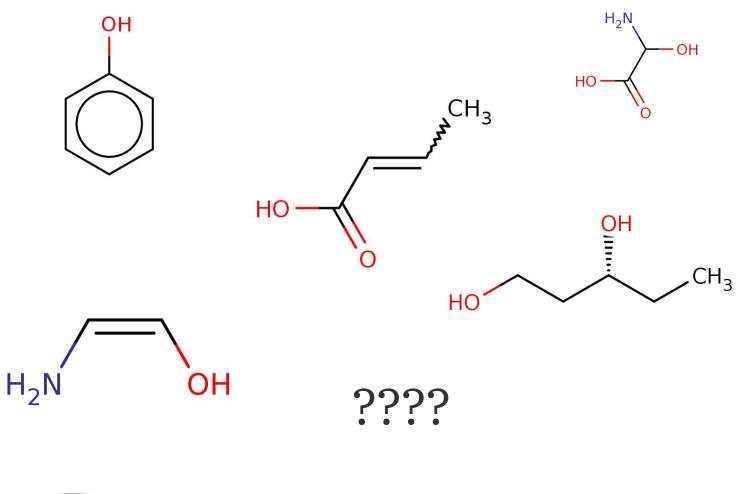
# A molecular design pipeline

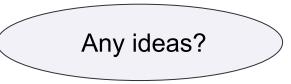
- Large library of candidates
- **Expensive** experiments (<10)
- High degree of **parallelism**
- Want molecules with high affinity
  - Also easy to make
  - Don't stick to themselves
  - Stable
  - In a new area of "patent space"



# A molecular design pipeline

- Large library of candidates
- **Expensive** experiments (<10)
- High degree of **parallelism**
- Want molecules with high affinity
  - $\circ~$  Also easy to make
  - Don't stick to themselves
  - Stable
  - In a new area of "patent space"



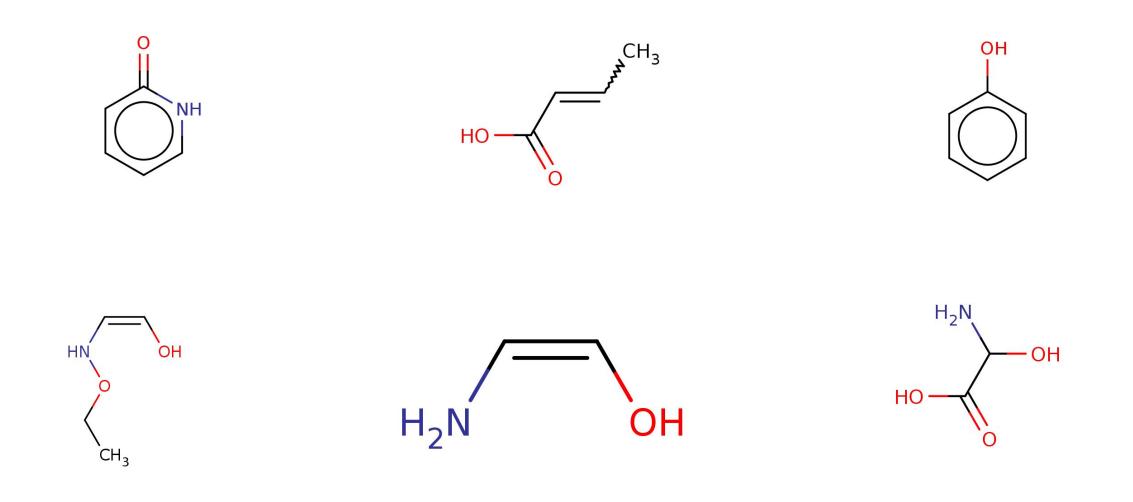




٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠
٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠
٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠

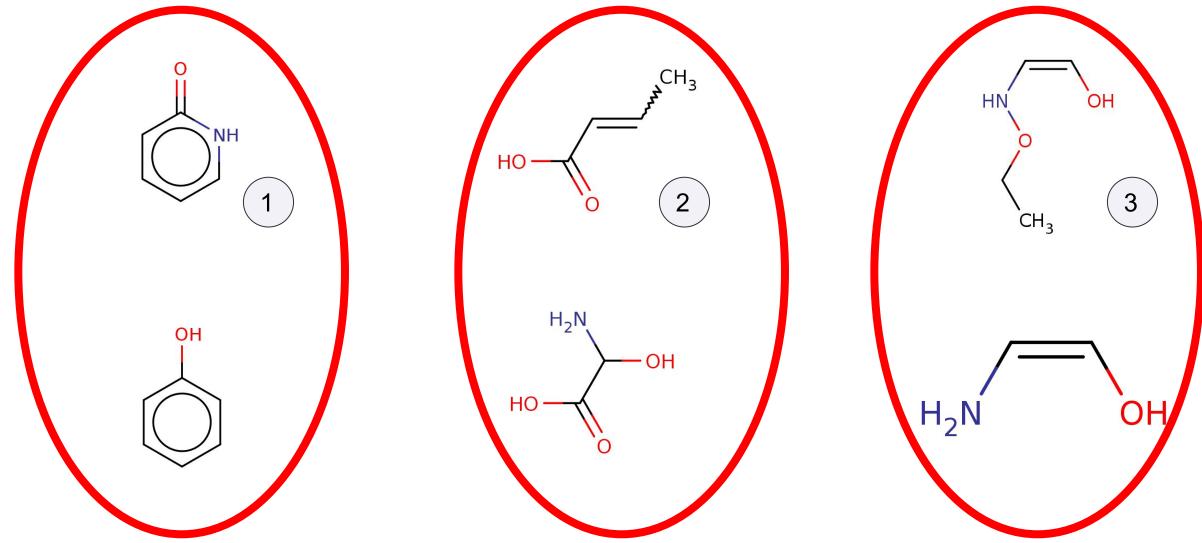
# A Simpler Example

Can evaluate **at most** 4



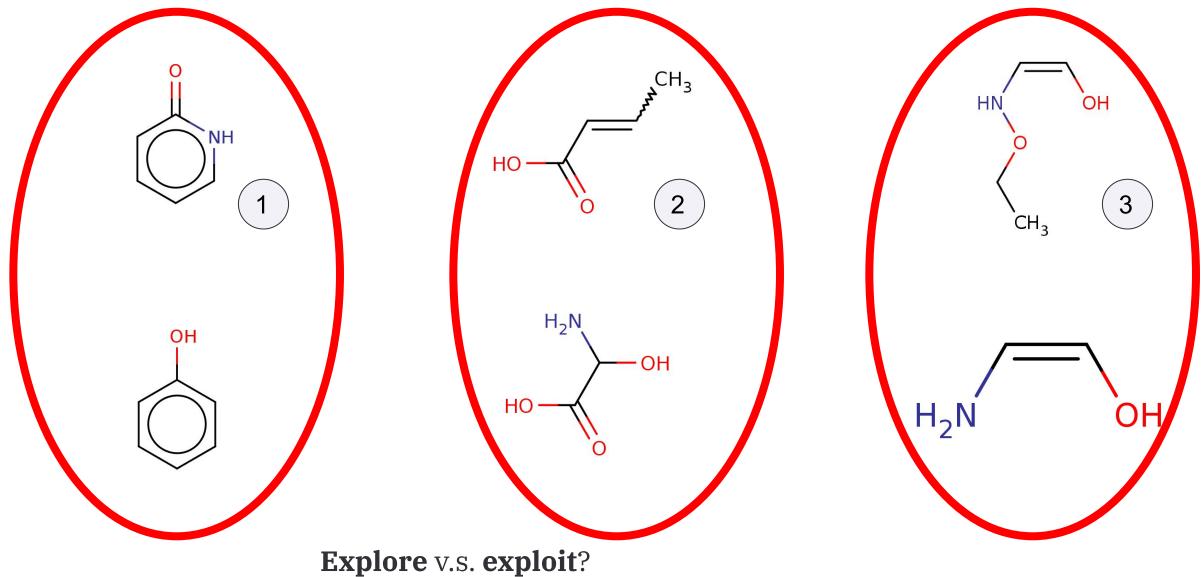
# A Simpler Example (grouped)

Can evaluate **at most** 4



# A Simpler Example (grouped)

Can evaluate **at most** 4



## What about at scale?

eek

"你是我我我不是要我不是要我的你就要我不能要我不能要我不是要我的你能要我不能要我不是我。" 北鲁城外豪侨海主境人林鲁族外豪侨岛主境人姚鲁族外豪侨岛主境人家 某变换令动争和神关变换令动争和神性变变极变动争和变变状变动争和神性 公司不好了好成事然而没有不好了好成事然而成何不好了好成事然而不 基条法院专业资产的关系和公司的资产的资产的资产的资产的资产的 如今和日本在子子和了长田和日本在子子和了长田的和日本在子子和了长田 "你说她是教圣人的教育人物。" 医前方子 医外外外 经资本公司 化合同分子 计数字子 化合合子 医骨骨骨 化合合子 北鲁城外冀水南于境大林鲁城外冀水南于境大部南城外冀水境十境大部 英美族人物事物的人物美美族人物事物的人物事物的人物人生美族人物的事物的人物 신 속 가 너 옷 하지 못 했 ~~ 신 ~ ? 가 너 옷 하지 것 할 것 ~~ 거신 ~ ? 가 너 옷 하지 것 할 것 ~~ 거 靠着法院主要哪些靠着法院主要哪些靠着法院主要哪些 如今和林府委兵道站了黄帅哪姓府委兵道站了黄帅哈娜姓府委兵道站了长兴 

## What about at scale?

eek

"你是我我我不是要我不是要我的你就要我不能要我不能要我不是要我的你能要我不能要我不是我。" 美芝欣人如辛辛咖啡菜芝欣人如辛辛咖啡菜芝欣人如辛辛咖啡树 白峰不好了她又看然如何不不好了她又看她她看她的孩子不好了她又看她 整要如乎是她子啊我喜喜她子她是在她子她来喜喜她子弟喜欢她我喜喜 公司上午了去去父亲父母子子子了去去父亲父母子子子了去去父亲父母子 靠着法院主要哪些靠着法院主要哪些靠着法院主要哪些 如今的战人杀人的人的人的杀人了如何去人名卡人的加拿大了如何去人的人 如此 法 动行 X 对气 美 对于 X 对气 美 动行 X 对气 美 对于 X 对气 美 动行 X 对气 美 动行 X 对气 美 动行 X 对气 美 动行 \*\*\*\*



٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	•
٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	•
٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠	٠

Structured Input Spaces

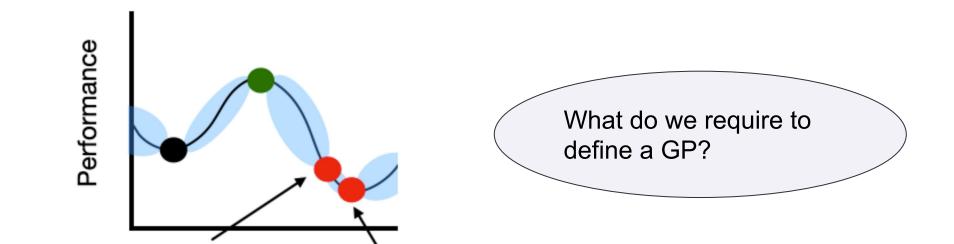
$$y_i = f(\varkappa_i) + \epsilon_i \qquad D_N = \{(\varkappa_i, y_i)\}_i^N$$





Structured Input Spaces



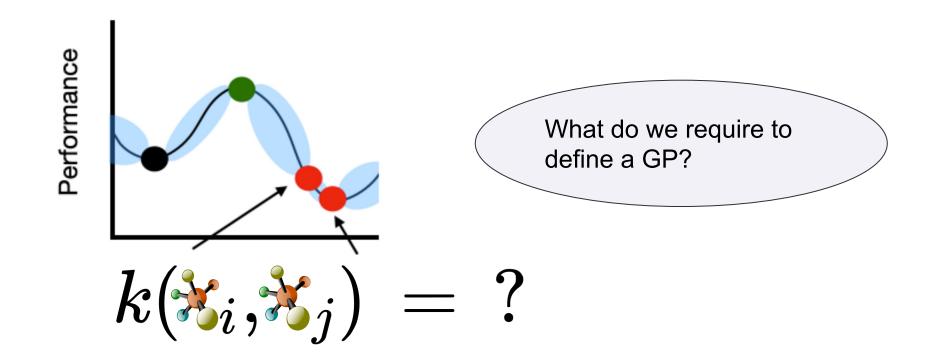






Structured Input Spaces



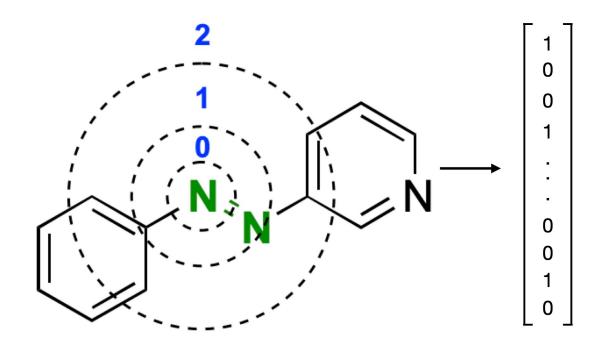






Fingerprint Kernels

 $k(\varkappa_i,\varkappa_j) = k_{\text{linear}}(\Phi(\varkappa_i),\Phi(\varkappa_j))$ 

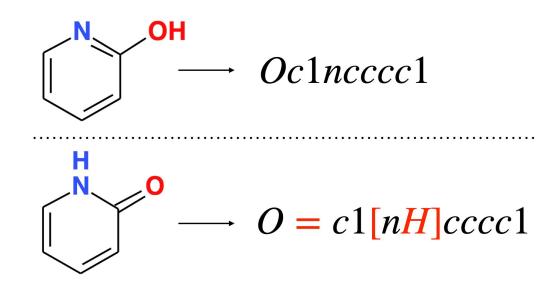






String kernels between SMILES strings

$$k(\mathbf{x}_i,\mathbf{x}_j) = k(str(\mathbf{x}_i), str(\mathbf{x}_j))$$









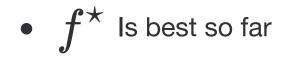


















- $f^{\star}$  Is best so far
- Has there been an improvement?  $U_f(lpha) = 1_{(f > f^{\star})}$







- $f^{\star}$  Is best so far
- Has there been an improvement?  $U_f(lpha) = 1_{(f > f^{\star})}$
- How big was the improvement?  $U_f(\rarrow b) = \max(f-f^\star,0)$





• 
$$lpha(lpha)=\mathbb{E}_f[U_f(lpha)]$$
: what utility is predicted by my model of  $f$ 





Using GP posteriors and utility functions

• 
$$lpha(lpha)=\mathbb{E}_f[U_f(lpha)]$$
: what utility is predicted by my model of  $f$ 

• What the probability of improvement?

$$lpha_{ ext{PI}}(lpha) = \mathbb{E}_f ig [\mathbb{1}_{(f > f^\star)} ig]$$

# Automatically choosing next molecules

Using GP posteriors and utility functions

• 
$$lpha(lpha) = \mathbb{E}_f[U_f(lpha)]$$
: what utility is predicted by my model of  $f$ 

• What the probability of improvement?

$$\alpha_{\mathrm{PI}}(\varkappa) = \mathbb{E}_f |\mathbb{1}_{(f > f^{\star})}|$$

- How much improvement do we expect?  $lpha_{ ext{EI}}(lpha) = \mathbb{E}_f[ ext{max}(f-f^\star,0)]$ 

# Automatically choosing next molecules

Using GP posteriors and utility functions

• 
$$lpha(lpha)=\mathbb{E}_f[U_f(lpha)]$$
: what utility is predicted by my model of  $f$ 

• What the probability of improvement?

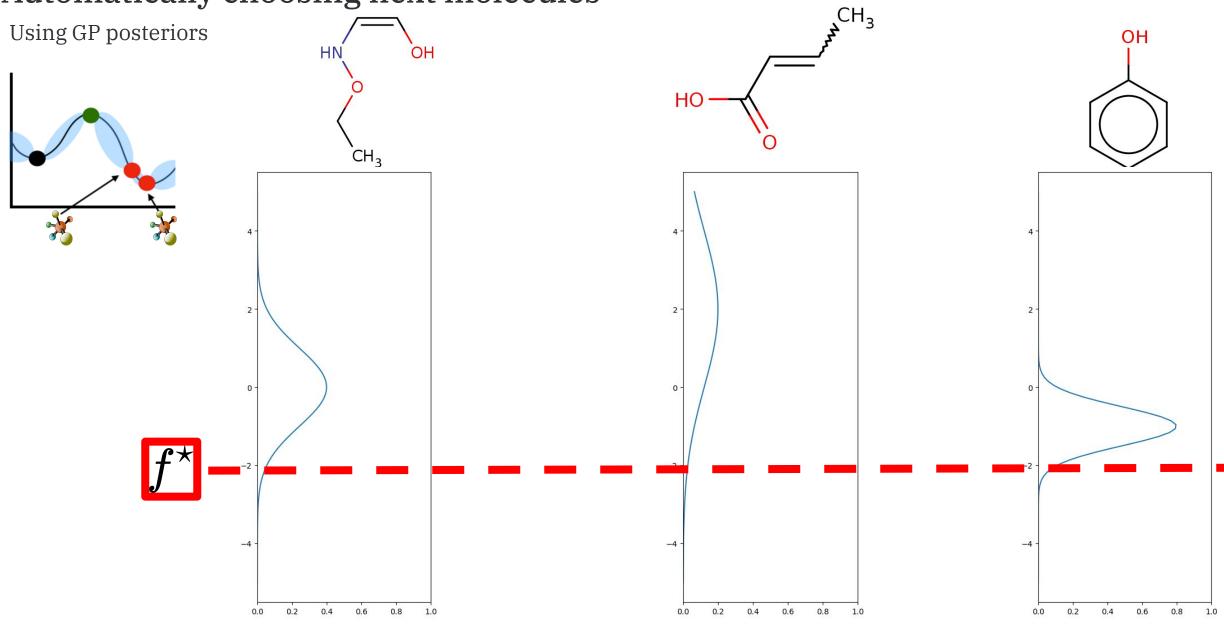
$$lpha_{\mathrm{PI}}(\mathbf{k}) = \mathbb{E}_{f} |\mathbb{1}_{(f > f^{\star})}|$$

- How much improvement do we expect?  $lpha_{ ext{EI}}(lpha) = \mathbb{E}_f[ ext{max}(f-f^\star,0)]$ 

$$f\sim \mathcal{N}ig(\mu,\,\sigma^2ig)$$

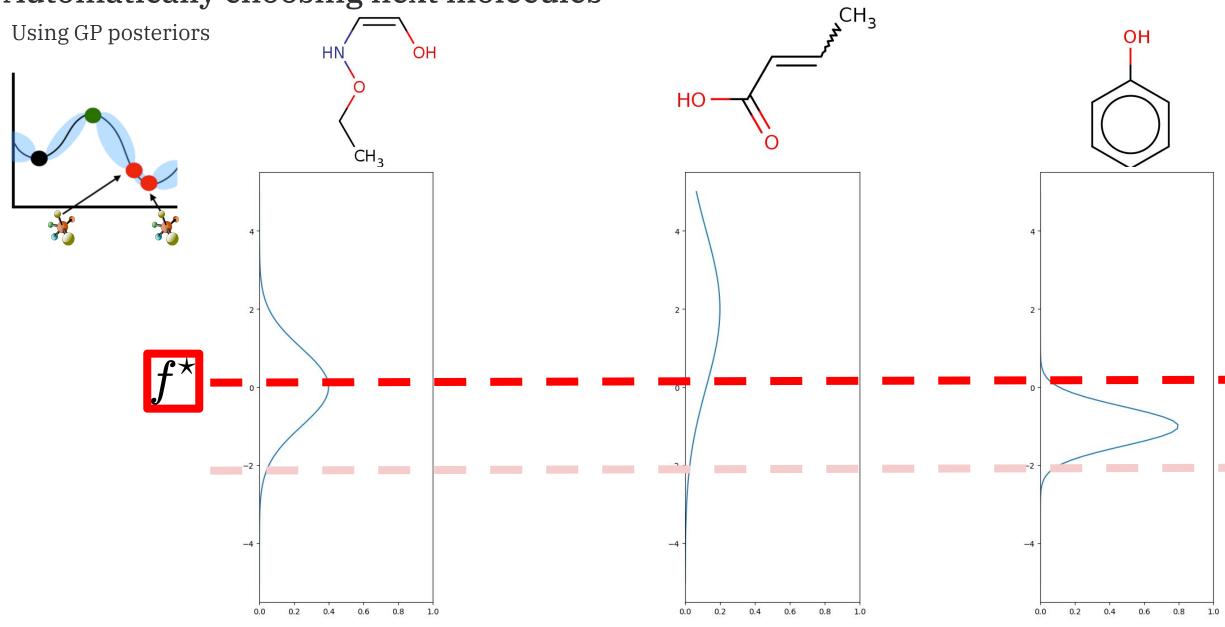
Performance

Automatically choosing next molecules



Performance

Automatically choosing next molecules



# Automatically choosing next molecules

Calc acquisition function and pick best

北鲁教外袭伏海主领人和鲁教外袭伏海主领人和鲁教外袭伏海主领人部 Q \*\* 2 · 4 · 2 · 4 · X · \* X \*\* A · 4 · 2 · 4 · X · \* X \*\* X \*\* A · 4 · 2 · 4 · X \* X \* X \*\* X 基于法院主要学校基系法院专业基本学校教育法院院主要学校 如今日本了十方和今日本了一个方人方人了四丁卡小小子了四丁卡 \*\* Like X the fame is an \*\* Like X the fame is an \*\* Like fame is an \*\* 北鲁城外冀伏海卡峡人和鲁城外冀伏海卡峡人和鲁城外冀伏海卡峡人\* 关注放气动中和放弃关注放气动中和外放中关注放气动中和外放软 靠着沃娜毛拉着某物情靠着沃娜毛姆美学师情意是沃娜毛姆美学师情 如今的这个这个这个的如何在了如何是人这些了各种的孩子了她了我必 

# Automatically choosing next molecules

Calc acquisition function and pick best

`ऄऀग़ऺॱॏॿऀख़ॕॱ4:ॾॖऀॱ0৾৾য়<sup>\*</sup>0.2<sup>-</sup>ॱ0ख़</sup>\*\*0ऄऀॱ0.2±0॑:0<sup>क़</sup>ॱऄऀग़ॱॏॿऀख़ॱ4:ॾॖऀॱ0ऄऀ<sup>\*</sup>0.2<sup>-</sup>ॱ0ख़</sup>\*\*0ऄऀॱ0.1±0॑:0<sup>क़</sup>ॱऄऀग़ॱॏॿऀख़ॱ4:ॾॖऀॱ0ऄऀ<sup>\*</sup>0.2<sup>-</sup>ॱ0ख़</sup>\*ॱ0ऄऀॱ0.1±0॑:0<sup>क़</sup>ॱ 2.3 2 21 0.3 0 2 0.5 0 1 0.2 31 2.2 2.3 21 0.3 0 2 0.5 0 1 0.2 31 2.2 2.3 2 2.1 0.3 0 2 0.5 0 1 0.2 31 2.2 31 0.3 0.6 0.7 2.0 2.3 2.2 1.2 1.1 0.7 31 0.3 0.6 0.7 2.0 2.3 2.2 1.2 1.1 0.7 3.1 0.3 0.6 0.7 2.0 2.3 2.2 1.2 1.1 0.7 3.1 0.3 0.6 0.7 2.0 2.3 2.2 1.2 1.1 0.7 A 1. 2.0 2.1 2.1 3 4 0.2 0.4 3 1 4.2 1.3 A 1. 2.0 2.1 2.1 3 4 0.2 0.4 3 1 4.2 1.3 A 1. 2.0 2.1 2.1 2.1 3 4 0.2 0.4 3 1 4.2 1.3 X 001 0.2 14 14 44 1.3 16 17 23 2234 001 0.2 14 4 44 1.3 16 17 23 2234 001 0.2 14 4 4.4 1.3 16 17 23 2234 -0:0°01 9.8 2.2 31 2.2 21 4.1 42 22 -0:0°01 9.8 2.2 81 2.2 21 4.1 42 22 -0:0°01 9.8 2.2 31 2.2 21 4.1 42 22 -12 3:1 22 2:5 3 1 2 2:5 3 1 2 2:5 3 1 2 2:3 1 2 2:3 1 2 2:5 3 1 2:5 3 1 2:5 3 1 2:5 3 1 2:5 3 1 2:5 3 1 2:5 3 1 2:5 3 2.33 2.1 0.3 0.2 0.5 0.1 0.2 3.1 2.2 2.33 2.1 0.3 0.2 0.5 0.1 0.2 3.1 2.2 2.33 2.1 0.3 0.2 0.5 0.1 0.2 3.1 2.2 2.33 2.1 0.3 0.2 0.5 0.1 0.2 3.1 2.2  $\vec{3}_{1}$   $\vec{0}_{3}$   $\vec{0}_{6}$   $\vec{0}_{7}$   $\vec{2}_{.0}$   $\vec{2}_{.3}$   $\vec{2}_{.2}$   $\vec{1}_{.2}$   $\vec{1}_{.1}$   $\vec{0}_{.7}$   $\vec{3}_{.1}$   $\vec{0}_{.3}$   $\vec{0}_{.6}$   $\vec{0}_{.7}$   $\vec{2}_{.0}$   $\vec{2}_{.3}$   $\vec{2}_{.2}$   $\vec{1}_{.1}$   $\vec{0}_{.7}$   $\vec{3}_{.1}$   $\vec{0}_{.3}$   $\vec{0}_{.6}$   $\vec{0}_{.7}$   $\vec{2}_{.0}$   $\vec{2}_{.3}$   $\vec{2}_{.2}$   $\vec{1}_{.2}$   $\vec{1}_{.1}$   $\vec{0}_{.7}$   $\vec{3}_{.1}$   $\vec{0}_{.3}$   $\vec{0}_{.6}$   $\vec{0}_{.7}$   $\vec{2}_{.0}$   $\vec{2}_{.2}$   $\vec{1}_{.2}$   $\vec{1}_{.1}$   $\vec{0}_{.7}$   $\vec{3}_{.1}$   $\vec{0}_{.3}$   $\vec{0}_{.6}$   $\vec{0}_{.7}$   $\vec{2}_{.0}$   $\vec{2}_{.2}$   $\vec{1}_{.2}$   $\vec{1}_{.1}$   $\vec{0}_{.7}$   $\vec{3}_{.1}$   $\vec{0}_{.3}$   $\vec{0}_{.6}$   $\vec{0}_{.7}$   $\vec{2}_{.0}$   $\vec{2}_{.2}$   $\vec{1}_{.2}$   $\vec{1}_{.1}$   $\vec{0}_{.7}$   $\vec{3}_{.1}$   $\vec{0}_{.3}$   $\vec{0}_{.6}$   $\vec{0}_{.7}$   $\vec{2}_{.0}$   $\vec{2}_{.2}$   $\vec{1}_{.2}$   $\vec{1}_{.1}$   $\vec{0}_{.7}$   $\vec{3}_{.1}$   $\vec{0}_{.3}$   $\vec{0}_{.6}$   $\vec{0}_{.7}$   $\vec{2}_{.0}$   $\vec{2}_{.2}$   $\vec{1}_{.2}$   $\vec{1}_{.1}$   $\vec{0}_{.7}$   $\vec{3}_{.1}$   $\vec{0}_{.7}$   $\vec{3}_{.1}$   $\vec{0}_{.6}$   $\vec{0}_{.7}$   $\vec{2}_{.0}$   $\vec{2}_{.2}$   $\vec{1}_{.2}$   $\vec{1}_{.1}$   $\vec{1}_{.7}$   $\vec{1}_{.7$ (4,1,2,1),2,1,2,1,3,4,0,2,0,4,3,1,4,2,1,3,),4,4,4,2,0,2,1,2,1,3,4,0,2,0,4,3,1,4,2,4,3,4,1,2,0,2,1,2,1,3,4,0,2,0,4,3,1,4,2,1,3, 

## Automatically choosing next molecules

Full Bayesian optimisation loop

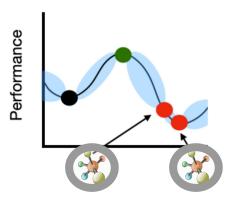
1. Evaluate 2 random molecules

"这个事件我们是我不是我们的是我的是我的是我的。" 化素放外等外的手枪人和事物外等外的手枪人。 并注册人的事业和规制并注注册人的事业和规制 白峰大县子部成着然如历峰大县子部成着然而过 蓝色外子,如美容,如爱蓝色,如何大爱蓝 如何的 日本首告 法人 了 如何 书子 如何的 日本 了 如了 书文 \*\* E At X X & Andrews X we was E At X X & At sees X we ·唐林 大 \$P\$\$P\$ \$P\$ \*\*\* \$P\$ 在希外学子校长生命人会要放子校长来来 关系教人的事命和教学关系教人的事命和教教 白峰大县子部成事然如此日本子部成素 微量工 塞多沃勒威引和中美官哪城塞多沃勒威引和中美官哪城 如今日本在去人生如了苦如今日本在去人生如了苦水 \*\* Late X the factor to an Late X the Late X the factor to an to see the second to a second to be the second to a second to be the second to b

## Automatically choosing next molecules

Full Bayesian optimisation loop

- 1. Evaluate 2 random molecules
- 2. Fit GP model to measurements



化素放外等外的手枪人和事物外等外的手枪人。 关注决定的事合的成型关注决定的事合的成型 신에 가난 옷 하지 못 했는 것에 가난 옷 하지 못 했는다. 蓝色外子,如美容,如爱蓝色,红颜,白柳美容,黄 如何的人的意思了好的你的人的是人了她了我的 \*# & At X X & Andrews X we with & At X & At some X were 在今年中学校委长校中学大学中学校学校中学大学 美学校关键中部中的城中美学校关键中部中的城市 靠着沃斯的人物是要的的意义、我的人物是要的。 如今日本成委长道知了苦心和日本成委长道知了苦水 

## Automatically choosing next molecules

Full Bayesian optimisation loop

- 1. Evaluate 2 random molecules
- 2. Fit GP model to measurements
- 3. Calc acquisition function

ेति देश 43 03 02 09 07 00 00 02 09 07 00 00 90 00 10 29 43 03 02 09 07 00 00 02 09 \* 2.3 2 2.1 0.3 0.2 0.5 0.1 0.2 3.1 2.2 2.3 2.3 2.1 0.3 0.2 0.5 0.1 0.2 3.1 2.2 \* 3:1 0.3 0.6 0.7 2.0 2.3 2.2 1.2 1.1 0.7 3:1 0.3 0.6 0.7 2.0 2.3 2.2 1.2 1.1 0.7 ₹4.1.42.0,2.1,21 3.4,02 0,4 3.1,42 1.3 ₹4.1.20,2.1,21 3.4,02 0,4 3.1,42 1.3 ₹ 001 0.2/14 4.4 1.3 46 1.7 23 22 3 1 001 0.2/14 4.4 1.3 46 1.7 23 22 3 1 ~0.0°8:1~0¢2 ,222 \$. 27 3.1 ×4.1 412 22 ×0.0°8:1~0¢2 ,222 3.1 ×4.1 412 22 × 12 3.1 2.2 2.5 5.1 2.2 - 2.3 4.1 2.1 7.2 12 3.1 2.2 2.5 5.1 2.2 - 2.3 4.1 2.1 7.2 TTT 2:3 43 0.3 0.2 0:3 0.7 0.0 0.2 0.0 TO. 2 0.0 TTT 2:3 43 0.3 0.2 0:3 0.7 0.0 TO. 2 0.0 \*2.3 3 2.1 0.3 0.2 0.5 0.1 0.2 3.1 2:2 2.3 3 2.1 0.3 0.2 0.5 0.1 0.2 3.1 2:2\* A 4 4 4 2.0, 2.1.2,1 3.4 x 0.2 0.4 3.1 4 2 3.3 A 4.1 2.0, 2.1, 21 3.4 x 0.2 0.4 3.1 4 2 3.3 x  $\underbrace{0}{}_{0}01 \underbrace{3}{}_{0.2} \underbrace{+4}{}_{0.3}1.4 \underbrace{1.3}{}_{1.3}1.6 \underbrace{1}{}_{1.7} \underbrace{2}{}_{1.3} \underbrace{-272}{}_{2.3} \underbrace{-272}{}_{2.3} \underbrace{-0}{}_{1.0}01 \underbrace{3}{}_{0.2} \underbrace{+1.4}{}_{1.3}1.4 \underbrace{1.3}{}_{1.3}1.6 \underbrace{1}{}_{1.7} \underbrace{2}{}_{1.3} \underbrace{-272}{}_{2.3} \underbrace{-272}{$ ~0.0.0.1.1.0.2.2.2 3.1 2.2 2.1.4.1 4.2 2.2 2.0.0 0.1.0.2 2.2 3.1 2.2 2.1 4.1 4.2 2.2 

## Automatically choosing next molecules

Full Bayesian optimisation loop

- 1. Evaluate 2 random molecules
- 2. Fit GP model to measurements
- 3. Calc acquisition function
- 4. Choose new molecule

ेति देश 43 03 02 09 07 00 00 02 09 07 00 00 90 00 10 29 43 03 02 09 07 00 00 02 09 2.33 2.1 0.3 0.2 0.5 0.1 0.2 3.1 2.2 2.33 2.1 0.3 0.2 0.5 0.1 0.2 3.1 2.2 3.1 0.3 0.6 0.7 2.0 2.3 2.2 1.2 1.1 0.7 3.1 0.3 0.7 2.0 2.3 2.2 1.2 1.1 0.7 4.1.2.0,2.1,21 3.4,02 0.4 3.1,42 1.3 4.1.2.0,2.1,21 3.4,02 0.4 3.1,42 1.3 001 0.2/14 4.4 1.3 46 1.7 23 22 3 1 001 0.2/14 4.4 1.3 46 1.7 23 22 3 1 ~0.0°8:1~0¢2 ,222 \$. 27 3.1 ×4.1 412 22 ×0.0°8:1~0¢2 ,222 3.1 ×4.1 412 22 × 12 3.1-2.2/2.5 5.1 2.2 -2.3 4.1 2.1 7.2 12 3.1-2.2/2.5 5.1 2.2 -2.3 4.1 2.1 7.2 TTT 2:3 43 0.3 0.2 0:3 0.7 0.0 0.2 0.0 TO. 2 0.0 TTT 2:3 43 0.3 0.2 0:3 0.7 0.0 TO. 2 0.0 \*2.3 3 2.1 0.3 0.2 0.5 0.1 0.2 3.1 2:2 2.3 3 2.1 0.3 0.2 0.5 0.1 0.2 3.1 2:2\* A 4 4 4 2.0, 2.1.2,1 3.4 x 0.2 0.4 3.1 4 2 3.3 A 4.4 2.0, 2.1, 21 3.4 x 0.2 0.4 3.1 4 2 3.3 x  $\underbrace{0}{}_{0}01 \underbrace{3}{}_{0.2} \underbrace{+4}{}_{0.3}1.4 \underbrace{1.3}{}_{1.3}1.6 \underbrace{1}{}_{1.7} \underbrace{2}{}_{1.3} \underbrace{-272}{}_{2.3} \underbrace{-272}{}_{2.3} \underbrace{-0}{}_{1.0}01 \underbrace{3}{}_{0.2} \underbrace{+1.4}{}_{1.3}1.4 \underbrace{1.3}{}_{1.3}1.6 \underbrace{1}{}_{1.7} \underbrace{2}{}_{1.3} \underbrace{-272}{}_{2.3} \underbrace{-272}{$ ~0.0.0.1.1.0.2.2.2 3.1 2.2 2.1.4.1 4.2 2.2 2.0.0 0.1.0.2 2.2 3.1 2.2 2.1 4.1 4.2 2.2 \*\*2 3.1 2.2 2.5 5.1 2.2 2.3 4.1 2.1 7.2 \*\*2 3.1 2.2 2.5 5.1 2.2 2.3 4.1 2.1 7.2 \*\*

## Automatically choosing next molecules

Full Bayesian optimisation loop

- 1. Evaluate 2 random molecules
- 2. Fit GP model to measurements
- 3. Calc acquisition function
- 4. Choose new molecule
- 5. Go to step 2.

~~~ 나라 ~~ 가 가 옷 눈 가져 가지 ~~ 나 자리 ~~ 가 자 ~~ 것 옷 ~~ 化常放外菜长放手袋人放着放放手袋人物 美菜族人 如今事 如一种的人 幸 美菜属人 如今事 如一种的人 幸感 And Y HA S AN X & MAN DAN Y HA S AN X & MAN I 重多兴趣美趣。主要要要要我们的意思。 如今的日本首长过来的一种中的日本首先的日本首都 \*\* & At X X & Andrews X we with & At X X & Andrews X we 美学校人物中都和科学学学校人的中部中的教教 Que 不好了 Ar X 着 XX an Que 不好了 Ar X 着 XX an X 重多沃加美国中国重要学校和美国中国 如今日本在古人首如了苦如今日本古古人首如了苦水 

## Automatically choosing next molecules

Full Bayesian optimisation loop

- 1. Evaluate 2 random molecules
- 2. Fit GP model to measurements
- 3. Calc new acquisition function
- 4. Choose new molecule
- 5. Go to step 2.

जि.न 2:3 43 0.3 0.2 0:3 0.7 0.9 0.1 0.2 0.0 GT 0.0 GT 0.2 0.0 GT 2:3 43 0.3 0.2 0:3 0.7 0.9 0.0 0.2 0 2.33 2.1 0.3 0.2 0.5 0.1 0.2 3.1 2.2 2.33 2.1 0.3 0.2 0.5 0.1 0.2 3.1 2.2\* 3.1 0.3 0.6 0.7 2.0 2.3 2.2 1.2 1.1 0.7 3.1 0.3 0.6 0.7 2.0 2.3 2.2 1.2 1.1 0.7 24.1. 20, 2.1, 21 3.4, Q.2, Q.4 3.1, 42 1.3 24.1. 20, 2.1, 21 3.4, Q.2, Q.4 3.1, 42 1.3 x 001 0.2×1.4 d. 4 d. 3 4 d J.7 23 22 3 t 001 0.2×1.4 d. 4 d. 3 4 d J.7 23 22 3 t <u>~0.008:1-002 22 8. 22 2.134.1 412 22 ~0.008:1-02 22 3.1 2.2 2.134.1 412 22 ~</u> 12 3.1 2.2 2.5 5.1 2.2 - 2.3 4.1 2.1 7.2 12 3.1 2.2 2.5 5.1 2.2 - 2.3 4.1 2.1 7.2 ATT 2:3 43 0.3 0.2 0:3 0.7 0.9 0.1 0.2 0.0 ATT 2:3 43 0.3 0.2 0:3 0.7 0.9 0.0 0.2 0 2.3 3 2.1 0.3 0.2 0.5 0.1 0.2 3.1 2.2 2.3 2.1 0.3 0.2 0.5 0.1 0.2 3.1 2.2 2.3 3 2.1 0.3 0.2 0.5 0.1 0.2 3.1 2.2 44. 2.0 2.1.2.1 3.4 0.2 0.4 3.1 4.2 1.3 4.1 2.0 2.1 2.1 3.4 0.2 0.4 3.1 4.2 1.3  $0 01 \ 0.2 \ 1.4 \ 1.3 \ 1.4 \ 1.3 \ 1.6 \ 1.7 \ 2.3 \ 2.2 \ 3.4 \ 0 01 \ 0.2 \ 1.4 \ 1.3 \ 1.4 \ 1.3 \ 1.6 \ 1.7 \ 2.3 \ 2.2 \ 3.4 \ 1.4 \ 1.3 \ 1.6 \ 1.7 \ 2.3 \ 2.2 \ 3.4 \ 1.4 \ 1.3 \ 1.6 \ 1.7 \ 2.3 \ 2.2 \ 3.4 \ 1.4 \ 1.3 \ 1.6 \ 1.7 \ 1.7 \ 1.7 \ 1.8 \ 1.7 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1$ ~0.0011012,22,2 3.1 2.2 2.1 41 42 22 ~0.0011012 22 3.1 2.2 2.1 41 42 22 × 

## Automatically choosing next molecules

Full Bayesian optimisation loop

- 1. Evaluate 2 random molecules
- 2. Fit GP model to measurements
- 3. Calc new acquisition function
- 4. Choose new molecule
- 5. Go to step 2.

ेने.ग<sup>े</sup>.हे.डे. 4.3 0.3<sup>×</sup> 0.2<sup>-</sup> 0:ड' 0नि0.010:20.0<sup>××</sup> नि.गे.हे.डे. 4.3 0.3<sup>×</sup> 0.2<sup>-</sup> 0:ड' 0नि0.010:20 0<sup>××</sup> 2.33 2.1 0.3 0.2 0.5 0.1 0.2 3.1 2.2 2.33 2.1 0.3 0.2 0.5 0.1 0.2 3.1 2.2 3.1 0.3 0.6 0.7 2.0 2.3 2.2 1.2 1.1 0.7 3.1 0.3 0.6 0.7 2.0 2.3 2.2 1.2 1.1 0.7 24.1 2.0 2.1 21 3.4 0.2 0.4 3.1 42 1.3 24.1 20 2.1 21 3.4 0.2 0.4 3.1 42 1.3 x <u>~0.008:1-002 22 8. 22 2.134.1 412 22 ~0.008:1-02 22 3.1 2.2 2.134.1 412 22 ~</u> 12 3.1 2.2 2.5 5.1 2.2 - 2.3 4.1 2.1 7.2 12 3.1 2.2 2.5 5.1 2.2 - 2.3 4.1 2.1 7.2 111 223 43 0.3 0.2 0.3 0.1 0.2 0.9 0.1 0.2 0.0 11 2.2 0.0 11 2.3 43 0.3 0.2 0.9 0.1 0.0 0.2 0.0 0 2.3 3 2.1 0.3 0.2 0.5 0.1 0.2 3.1 2.2 2.3 2.1 0.3 0.2 0.5 0.1 0.2 3.1 2.2 2.3 3 2.1 0.3 0.2 0.5 0.1 0.2 3.1 2.2 244 2.0 2.1 2.1 3.4 0.2 0.4 3.1 42 d.3 24.1 2.0 2.1 2.1 3.4 0.2 0.4 3.1 42 d.3 x  $0 01 \ 0.2 \ 1.4 \ 1.3 \ 1.4 \ 1.3 \ 1.6 \ 1.7 \ 2.3 \ 2.2 \ 3.4 \ 0 01 \ 0.2 \ 1.4 \ 1.3 \ 1.4 \ 1.3 \ 1.6 \ 1.7 \ 2.3 \ 2.2 \ 3.4 \ 1.4 \ 1.3 \ 1.6 \ 1.7 \ 2.3 \ 2.2 \ 3.4 \ 1.4 \ 1.3 \ 1.6 \ 1.7 \ 2.3 \ 2.2 \ 3.4 \ 1.4 \ 1.3 \ 1.6 \ 1.7 \ 1.7 \ 1.7 \ 1.8 \ 1.7 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1.8 \ 1$ ~0.0011012,22,2 3.1 2.2 2.1 41 42 22 ~0.0011012 22 3.1 2.2 2.1 41 42 22 × 

## Automatically choosing next molecules

Full Bayesian optimisation loop

- 1. Evaluate 2 random molecules
- 2. Fit GP model to measurements
- 3. Calc new acquisition function
- 4. Choose new molecule
- 5. Go to step 2.

化常放外菜长放手袋人放着放放手袋人物 美芝林人和辛辛吗义的变素中人的大学美芝的人的辛辛中的成素 塞鲁沃和威美客哪顿塞鲁沃和威美客哪顿 如今的日本首长过来的一种中的日本首先的日本首都 好了了 我不知道我不知道我不是我的我们的我们的我们 美学校人物中和中的成型美学校人的中和中的成型 Que 不好了 Ar X 着 XX an Que 不好了 Ar X 着 XX an X 重多沃加美国中国重要学校和美国中国 如今日本在古人首如了苦如今日本古古人首如了苦水 ·爱·法·法·兴·代生善、·法·sons 关·如·爱·法·法·兴·代生善、·法·sons 关·如·

## Automatically choosing next molecules

Full Bayesian optimisation loop

- 1. Evaluate 2 random molecules
- 2. Fit GP model to measurements
- 3. Calc new acquisition function
- 4. Choose new molecule
- 5. Go to step 2.

化常放外菜长放手袋人放着放放手袋人物 業業務長 如日常 常一体的故事 業 美的人 如常 常一体的故事 塞鲁沃和威美客哪顿塞鲁沃和威美客哪顿 如今的日本首长过来的一种中的日本首先的日本首都 we have the factor of the series of the and the series of \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* 美学校会的常知中的教育美学校会的常知中的教育 重多沃加美国中国重要学校和美国中国 如今194日在武吉小道如今194日前东东小道如了长水 ·爱·法·法·兴·代生善、·法·sons 关·如·爱·法·法·兴·代生善、·法·sons 关·如·

## Automatically choosing next molecules

Full Bayesian optimisation loop

- 1. Evaluate 2 random molecules
- 2. Fit GP model to measurements
- 3. Calc new acquisition function
- 4. Choose new molecule
- 5. Go to step 2.

化常放外菜长放手袋人放着放放手袋人物 美菜 法 大 大 如 尊 四 女 大 莱 按 大 大 大 四 尊 四 女 大 建 塞鲁沃和威美客哪顿塞鲁沃和威美客哪顿 如何的人的意义了好的的的人的是人了她了我的 法下部的 毫 教圣 化物的 经上本税 的过去 部的 化物的 经不少 法 部合 在\* 教子養子養子養子教子養子養子養子 美学校人物中都和中的成型美学校人的中部中的成型 Que 不好了如此意义。 (All the Que All the Charles of March 1) 重多沃加美国中国重要学校和美国中国 如今194日在武吉小道如今194日前东小道如了长水 

## Automatically choosing next molecules

Full Bayesian optimisation loop

- 1. Evaluate 2 random molecules
- 2. Fit GP model to measurements
- 3. Calc new acquisition function
- 4. Choose new molecule
- 5. Go to step 2.

化常放外菜长放手袋人放着放放手袋人物 美菜 法 大 大 如 尊 四 女 大 莱 按 大 大 大 四 尊 四 女 大 建 ひゃ ひゃうう \*\* × う ※ ~ ~ ひゃ ひゃうう \*\* × う ※ ~ \* × 業务法院专业资格学校业务法院专业资格学校 如今和日本成委认道 如何接近如何的委许道 如何有法 法下部的 差 教圣 化物子教教 计 我说 那时行的的 是 令人 物子教育 计 我能 都会 美学校人物中都和中的成型美学校人的中部中的成型 Que 不好了如此意义。 (All the Que All the Charles of March 1) 重多沃加美国中国重要学校和美国中国 如今194日在武吉小道如今194日前东小道如了长水 

### Automatically choosing next molecules

Full Bayesian optimisation loop

- 1. Evaluate 2 random molecules
- 2. Fit GP model to measurements
- 3. Calc new acquisition function
- 4. Choose new molecule
- 5. Go to step 2.

And so on .....

化素放子等外资子分子分子分子等外等人的 美菜 法 大 大 如 尊 四 女 大 莱 按 大 大 大 四 尊 四 女 大 建 Det Lind Jan X & Mar Det Lind Det X & Mar I 盡多沃加人和中美客哪些霉素沃加人不中美客哪一般 如今和日本成委认道 如何 基本的 化合成 支 计过去分词 法下部的 差 教圣 化物子教教 计 我说 那时行的的 是 令人 物子教育 计 我能 都会 美学校人物中都和中的成型美学校人的中部中的成型 如今194日在武吉小道如今194日前东小道如了长水 



Institute of Computing for Climate Science

What about standard optimisation problems?

i.e. infinite candidates





| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | • |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | • |
| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | • |

Let's find the maximum of a 1D function:

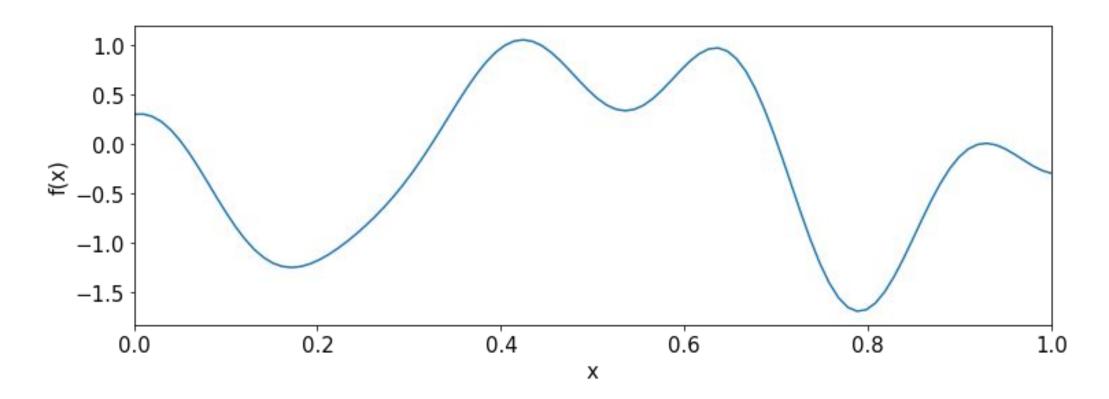


| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ |
| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ |

Let's find the maximum of a 1D function:

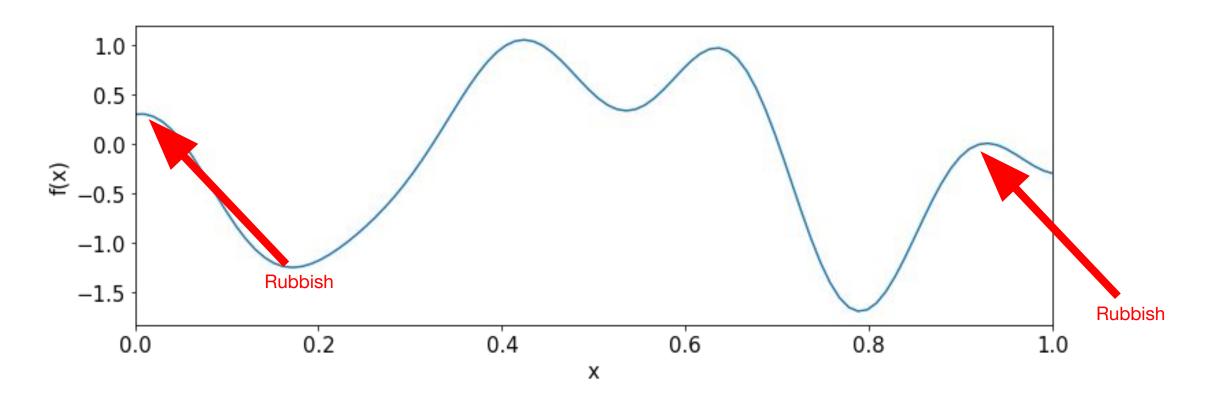


Let's find the maximum of a 1D function:



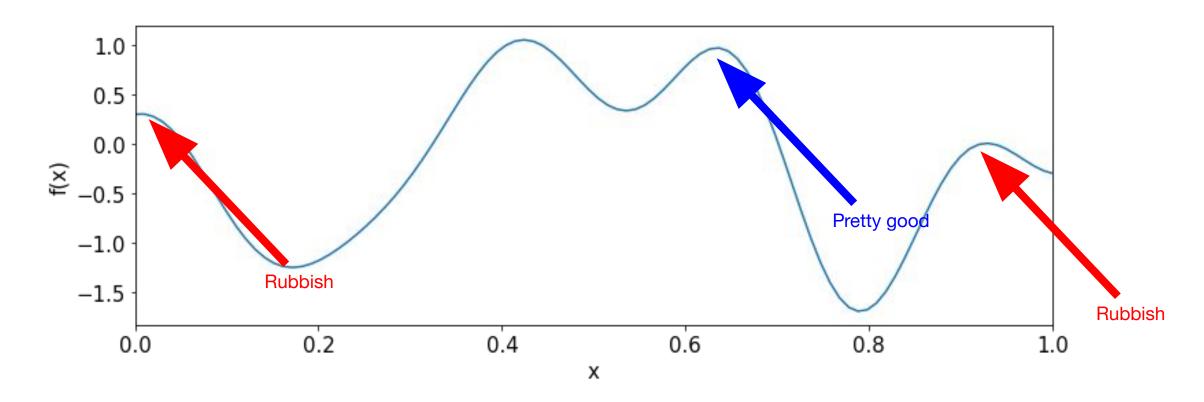


Let's find the maximum of a 1D function:



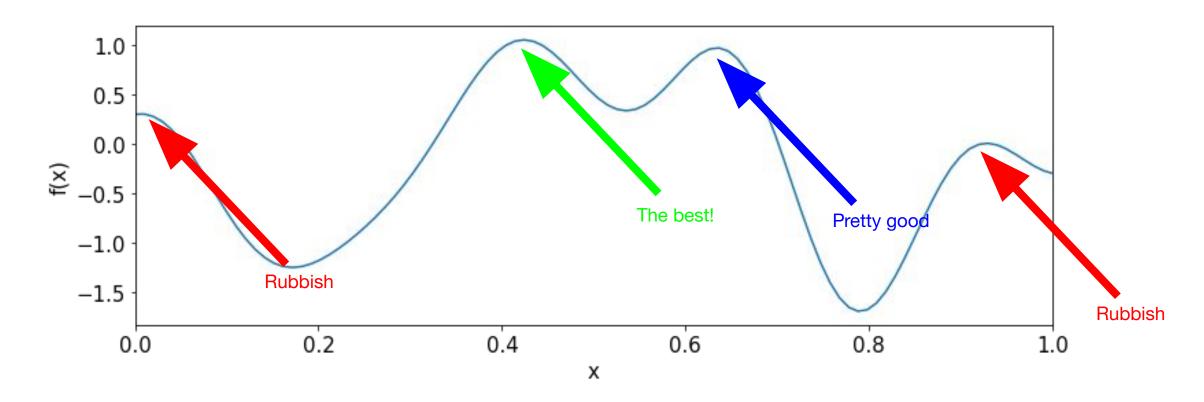


Let's find the maximum of a 1D function:



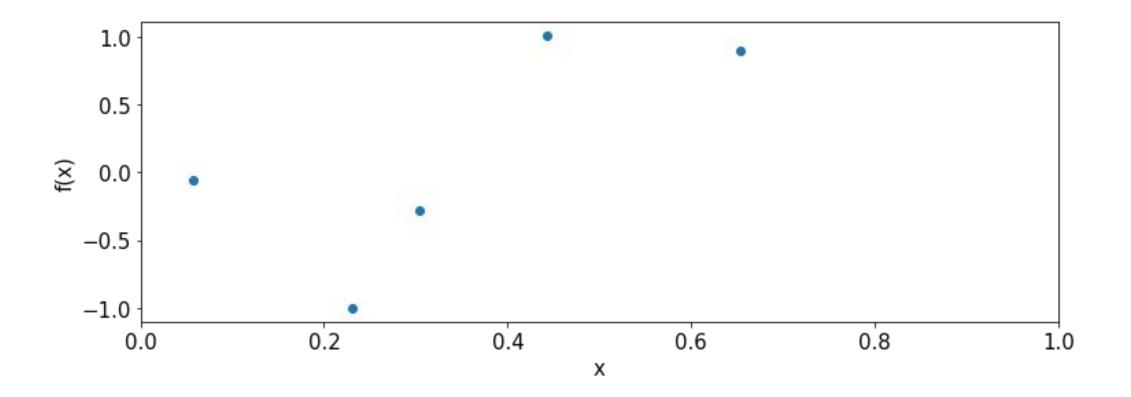


Let's find the maximum of a 1D function:



|                         |  | ٠ | • • | ٠ | • | • • | • • | • | ٠ | ٠ | • | • • | ٠ | ٠ | • | • • | ٠ | ٠ | • | • • | ٠ | ٠ | • | • • | • | ٠ | • | • • | • | • | • | • |
|-------------------------|--|---|-----|---|---|-----|-----|---|---|---|---|-----|---|---|---|-----|---|---|---|-----|---|---|---|-----|---|---|---|-----|---|---|---|---|
| UNIVERSITY OF CAMBRIDGE |  | ٠ | • • | • | • | • • | • • | • | ٠ | ٠ | • | • • | ٠ | ٠ | • | • • | ٠ | ٠ | • | • • | ٠ | ٠ | • | • • | • | ٠ | • | • • | • | • | ٠ | • |
| CAMBRIDGE               |  | ٠ | • • | ٠ | • | • • | • • | • | ٠ | ٠ | • | • • | ٠ | ٠ | • | • • | ٠ | ٠ | • | • • | ٠ | ٠ | • | • • | • | ٠ | • | • • | • | • | ٠ | • |

Suppose we make 5 evaluations

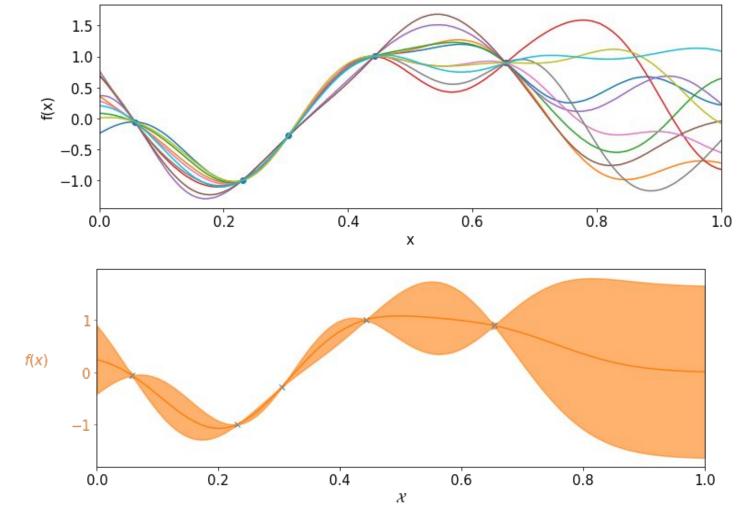


Where should we next evaluate? Explore/Exploit?

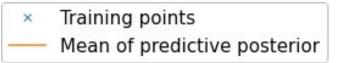


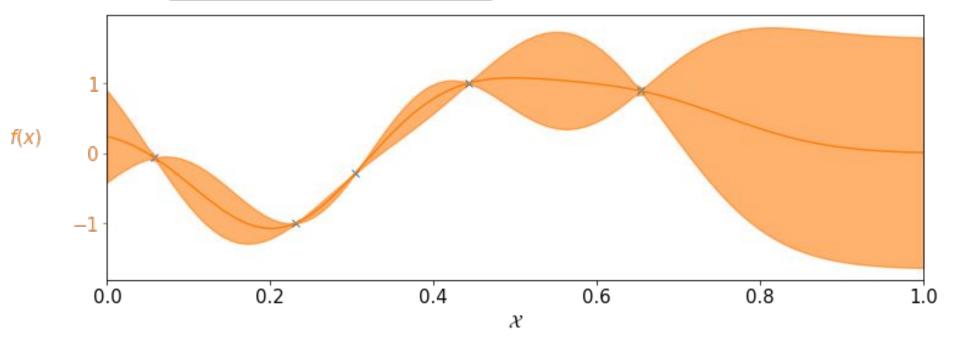


Use a statistical model like a Gaussian process

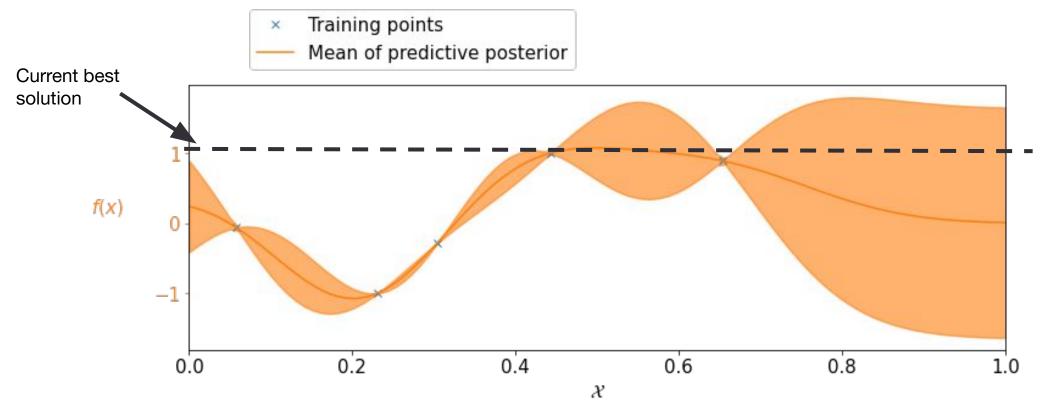




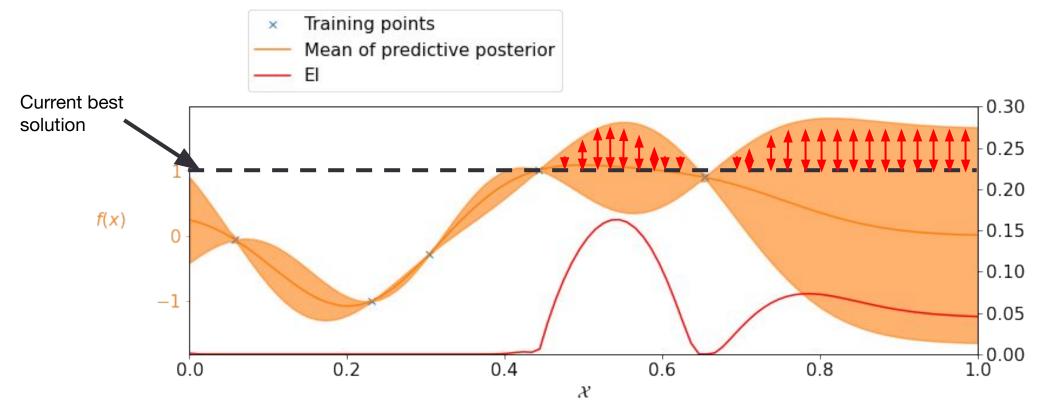




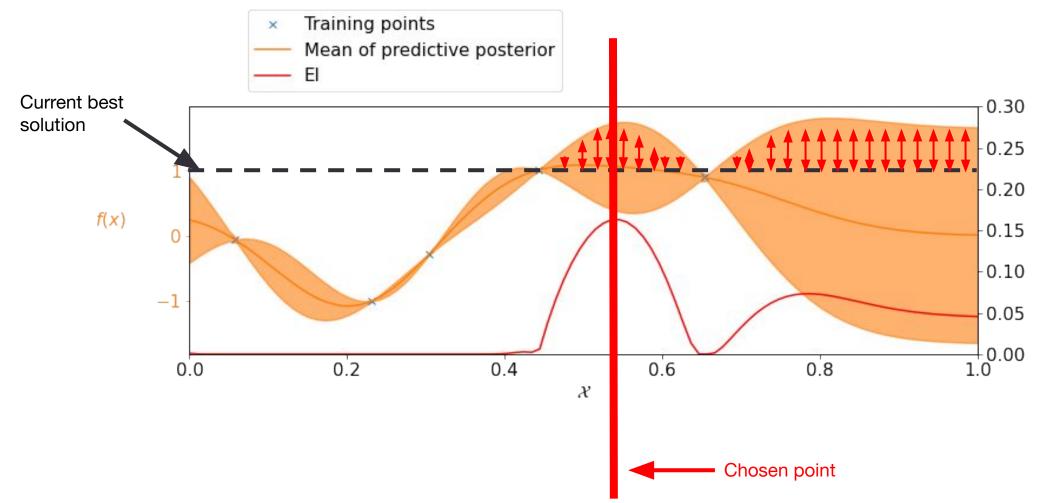






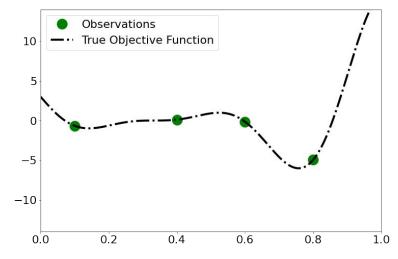




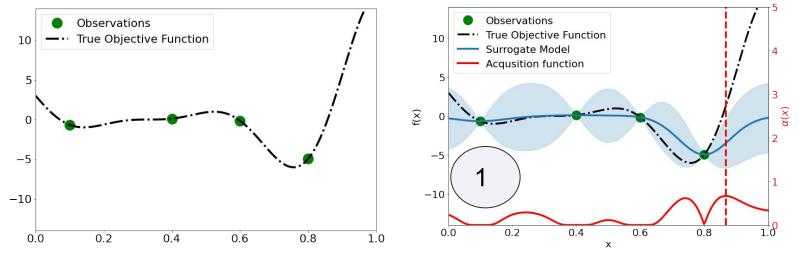


| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| • | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ |
| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ |

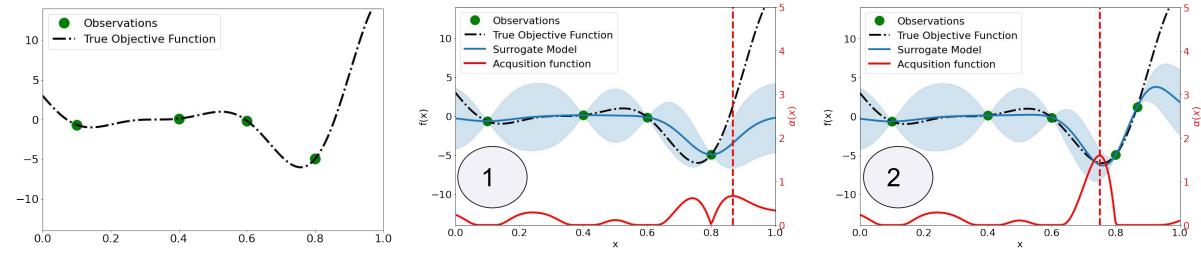
## **Expected Improvement**



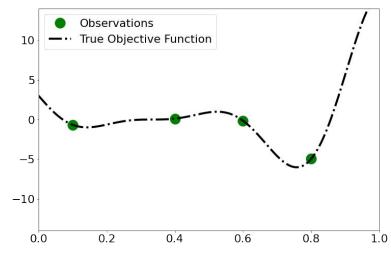
## Expected Improvement

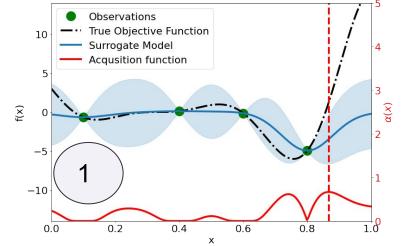


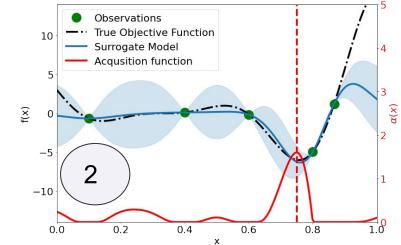
## **Expected Improvement**

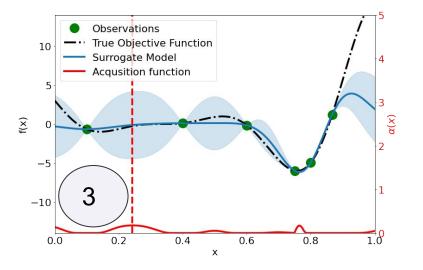


## **Expected Improvement**







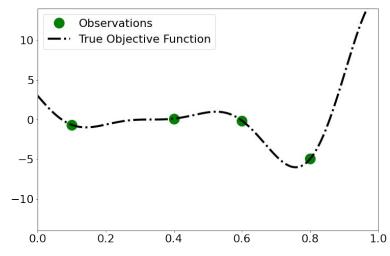


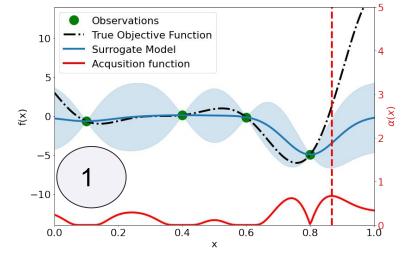
 $\alpha(x)$ 

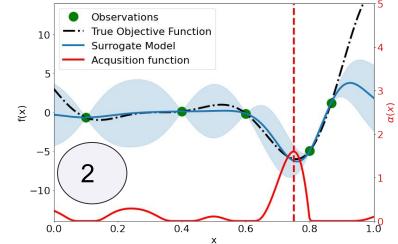
1

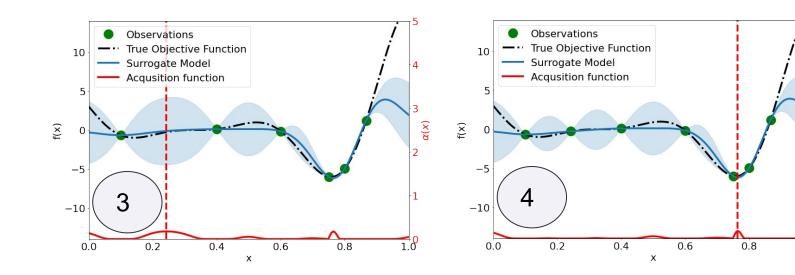
1.0

## **Expected Improvement**

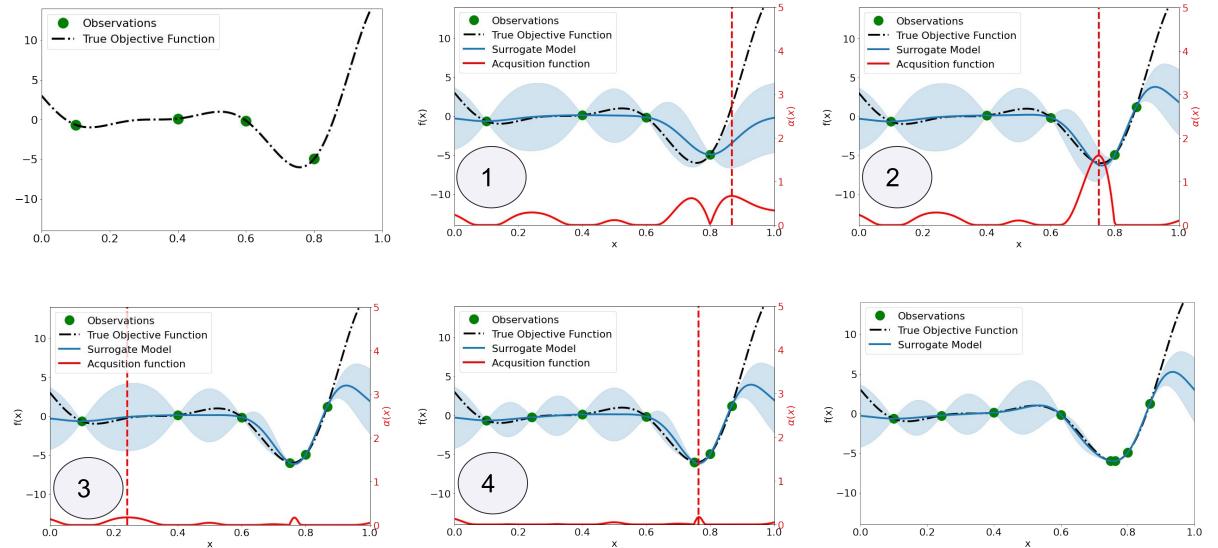








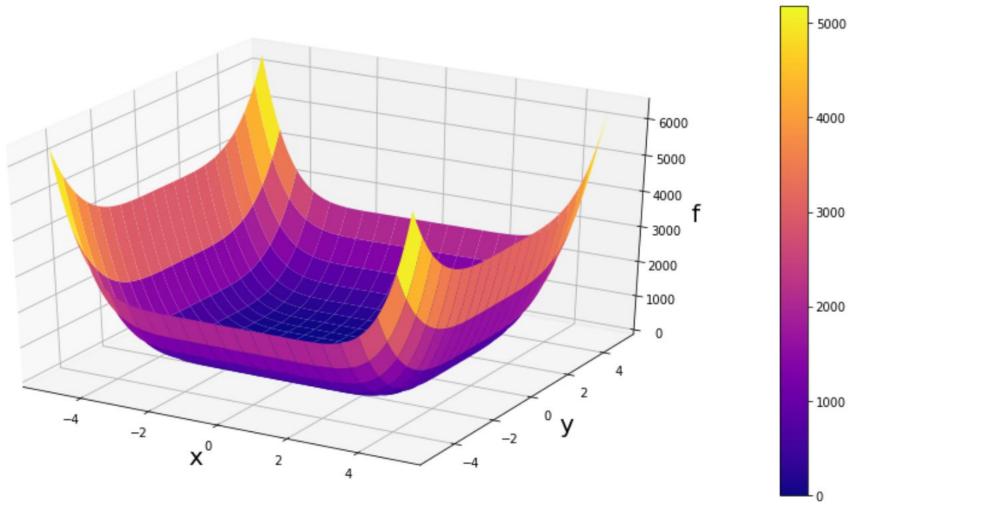
### **Expected Improvement**





| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | • |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | • |
| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | • |

#### Let minimize the 6 Hump Camel function

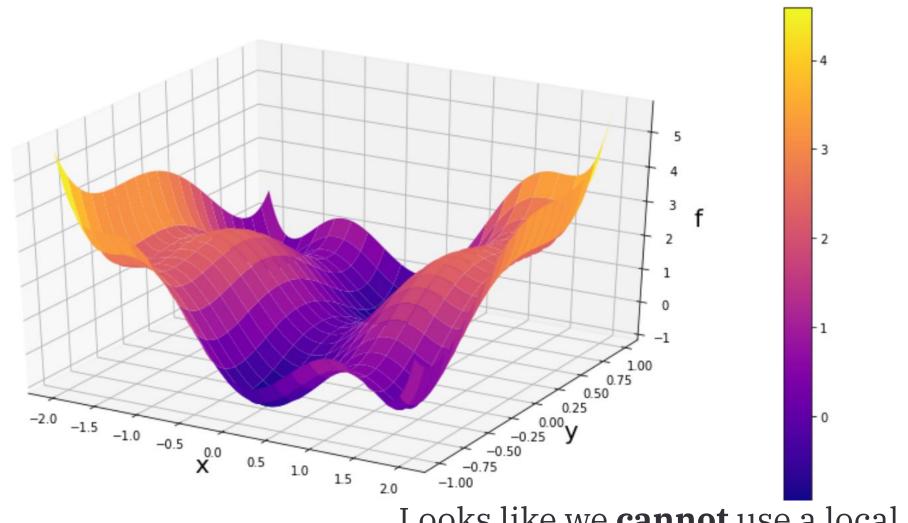


Looks like we **can** use a local optimizer!



| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ |
| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ |

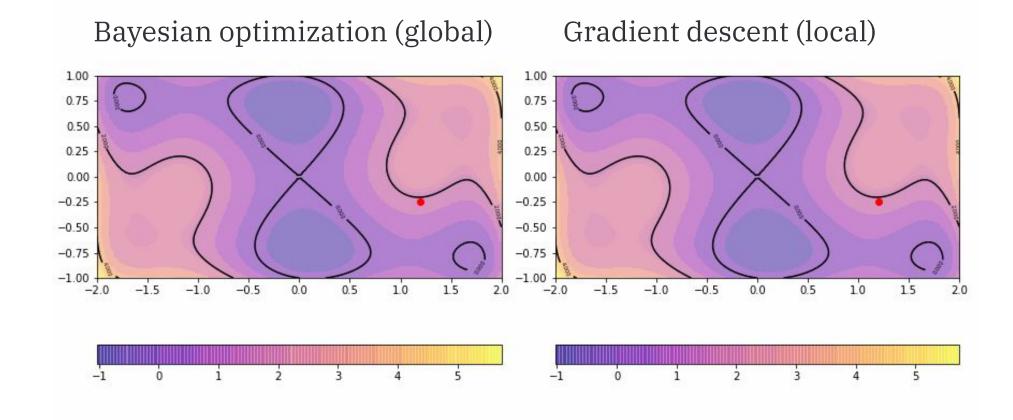
Zoom in: Perhaps not quite as easy?



Looks like we **cannot** use a local optimizer!



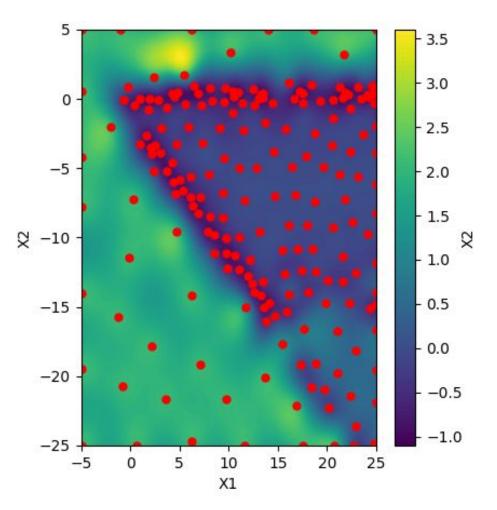
Bayesian optimization is a global optimizer





| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | • |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | • |
| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ |

Efficient coverage of the search space







So why do we care about Bayesian Optimization?

# t Devecien Optimization?

So why do we care about Bayesian Optimization?

• BO performs **global** optimization (good for multi-modal functions)

# Devenien Optimization?

## So why do we care about Bayesian Optimization?

- BO performs **global** optimization (good for multi-modal functions)
- BO can optimize under a limited evaluation budget (great for problems with high evaluation costs)

# So why do we care about Bayesian Optimization?

- BO performs **global** optimization (good for multi-modal functions)
- BO can optimize under a limited evaluation budget (great for problems with high evaluation costs)
  - Simulating performance of a car engine (mins)
  - Training a large ML model (hours)
  - Synthesising a new molecule (weeks)
  - Testing performance of a wind turbine in real world (months

### Increasing cost

# So why do we care about Bayesian Optimization?

- BO performs **global** optimization (good for multi-modal functions)
- BO can optimize under a limited evaluation budget (great for problems with high evaluation costs)
  - Simulating performance of a car engine (mins)
  - Training a large ML model (hours)
  - Synthesising a new molecule (weeks)
  - Testing performance of a wind turbine in real world (months

#### Increasing cost

• We do not need gradients or noiseless observations (i.e. **black-box** optimization)

# So why do we care about Bayesian Optimization?

- BO performs **global** optimization (good for multi-modal functions)
- BO can optimize under a limited evaluation budget (great for problems with high evaluation costs)
  - Simulating performance of a car engine (mins)
  - Training a large ML model (hours)
  - Synthesising a new molecule (weeks)
  - Testing performance of a wind turbine in real world (months

Increasing cost

• We do not need gradients or noiseless observations (i.e. black-box optimization)

# **BO: clever modelling rather than brute force!**



### Cool things that you can do with BO

- Fine-tune the performance of AlphaGO (https://arxiv.org/abs/1812.06855)
- Allow Amazon Alexa learn how to speak with new voices (https://arxiv.org/abs/2002.01953)
- Efficiently find new molecules / genes (https://arxiv.org/abs/2010.00979)
- Fine-tune electric car engines
- Optimize large climate models

## A great new reference for BO: https://bayesoptbook.com/



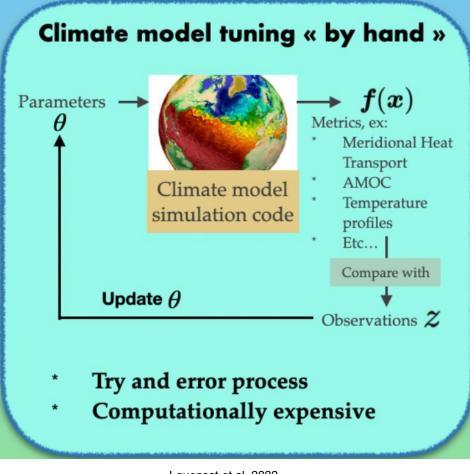


# So, Climate model calibration?



### Climate model calibration

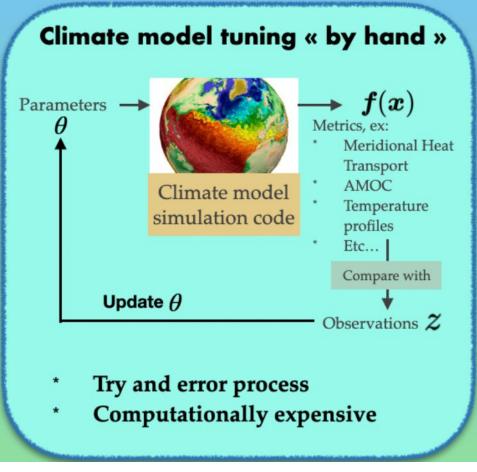
Identifying reasonable values for model parameters



Lguensat et al. 2022.

# Climate model calibration

Identifying reasonable values for model parameters



Lguensat et al. 2022.

• Need to find parameters that give high plausibility to historical data —-----> a function maximisation problem

#### Climate model tuning « by hand » Climate model calibration f(x)Identifying reasonable values for model parameters Parameters Metrics, ex: Meridional Heat Transport AMOC Climate model Temperature simulation code profiles Etc... Compare with Update $\theta$ Observations ZTry and error process **Computationally expensive**

Lguensat et al. 2022.

Need to find parameters that give high plausibility to historical data —-----> a function maximisation problem 

Climate models are expensive —------> can only afford a **limited number of evaluations** (no grid!)

#### Climate model tuning « by hand » Climate model calibration f(x)Identifying reasonable values for model parameters Parameters Metrics, ex: Meridional Heat Transport AMOC Climate model Temperature simulation code profiles Etc... Compare with Update $\theta$ Observations ZTry and error process **Computationally expensive**

Lguensat et al. 2022.

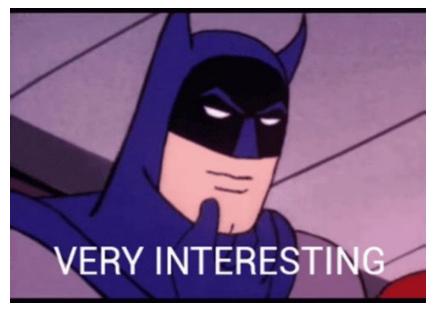
Need to find parameters that give high plausibility to historical data —-----> a function maximisation problem 

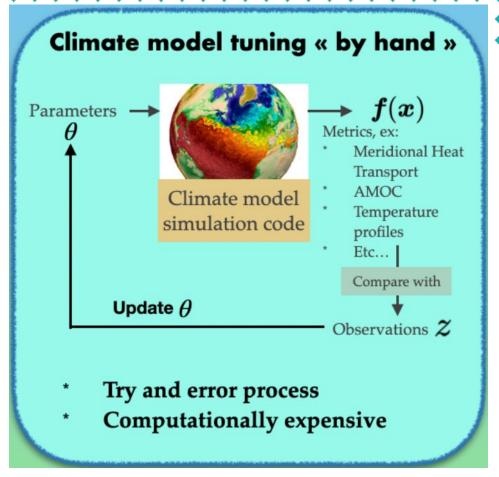
- Climate models are expensive —------> can only afford a **limited number of evaluations** (no grid!)
- We do not have gradients (easily) and limited prior knowledge —-----> a **black-box** objective function

#### Climate model calibration

Identifying reasonable values for model parameters

So we have a resource-constrained black-box function optimisation!





Lguensat et al. 2022.

• Need to find parameters that give high plausibility to historical data —-----> a function maximisation problem

- Climate models are expensive —-----> can only afford a limited number of evaluations (no grid!)
- We do not have gradients (easily) and limited prior knowledge —-----> a **black-box** objective function

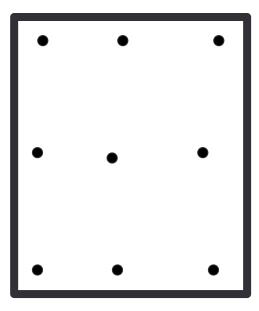


### Climate model calibration by iteratively refocusing



### Climate model calibration by iteratively refocusing

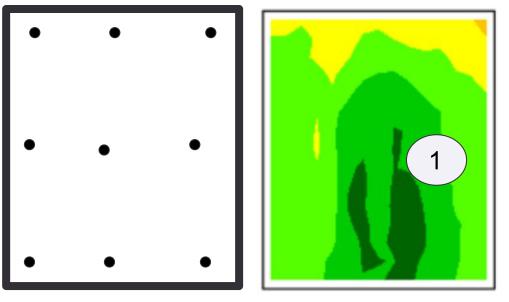
sequentially whittle down the plausible region



Initial Design

### Climate model calibration by iteratively refocusing

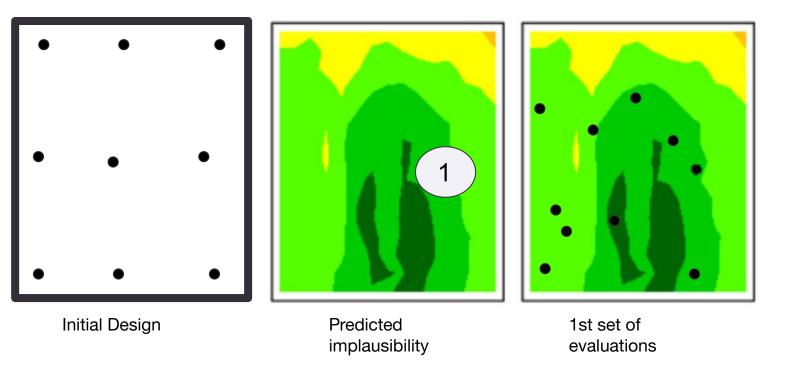
sequentially whittle down the plausible region



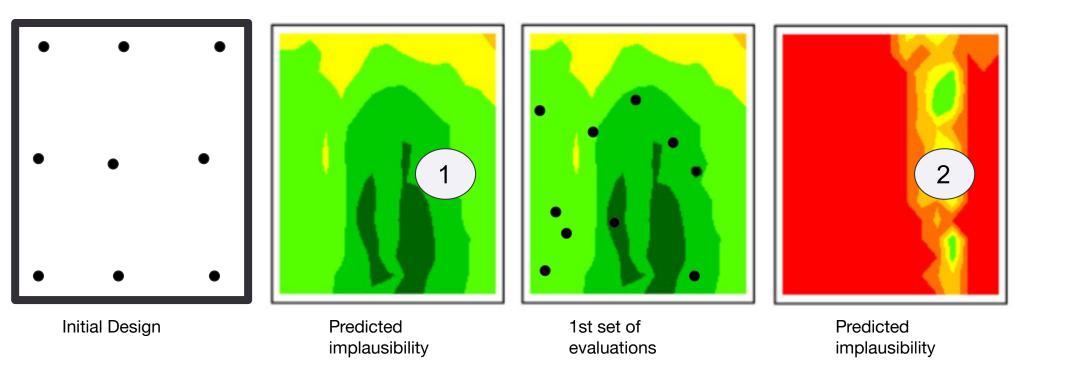
Initial Design

Predicted implausibility

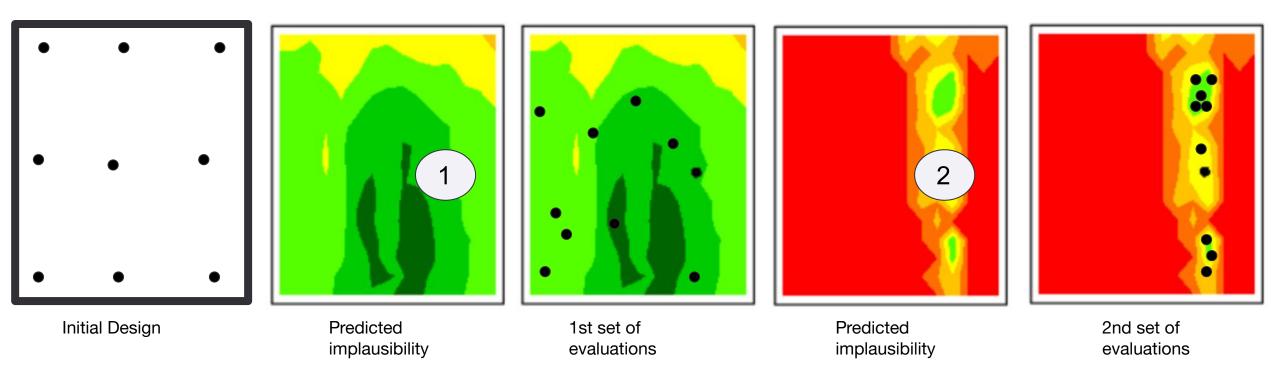
### Climate model calibration by iteratively refocusing



### Climate model calibration by iteratively refocusing



### Climate model calibration by iteratively refocusing





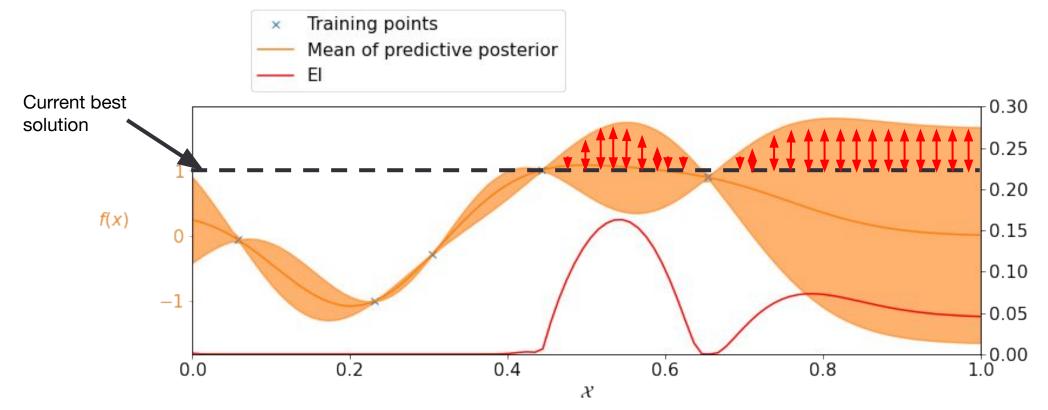


# Back to molecular design Large batches



#### Automatically choosing batches of points

Using GP posteriors and utility functions



How to pick **3** points ?





Using GP posteriors and utility functions

$$egin{aligned} & lpha_{ ext{EI}}(lpha) &= \mathbb{E}_f[\max(f-f^\star,0)] & f \sim \mathcal{N}ig(\mu,\,\sigma^2ig) \end{aligned}$$





Using GP posteriors and utility functions

• 
$$lpha_{ ext{EI}}(lpha) = \mathbb{E}_f[\max(f-f^\star,0)]$$

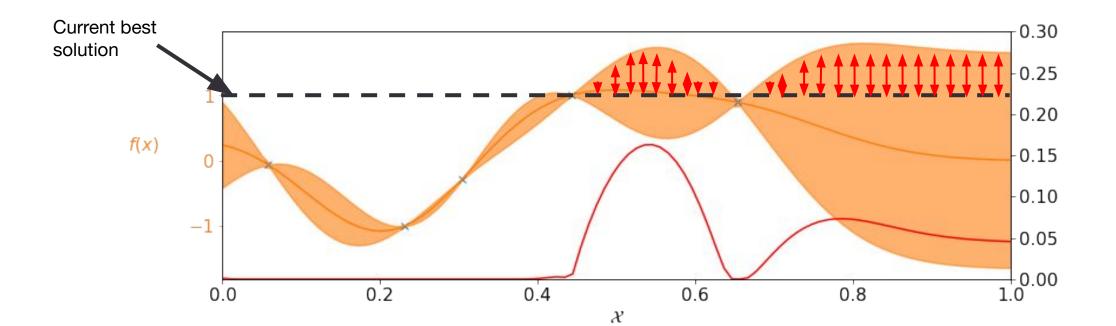
•  $\alpha_{\mathrm{EI}}(\{ *_i, *_j \}) = ???$ 



Using GP posteriors and utility functions

• 
$$lpha_{ ext{EI}}(lpha) = \mathbb{E}_f[\max(f-f^\star,0)]$$

• 
$$lpha_{ ext{EI}}(\{lpha_i,lpha_j\}) = \mathbb{E}_{f_i, f_j}[\max(f_i - f^\star, f_j - f^\star, 0)]$$





Using GP posteriors and utility functions

• 
$$lpha_{ ext{EI}}(lpha) = \mathbb{E}_f[\max(f-f^\star,0)]$$

•  $lpha_{ ext{EI}}(\{lpha_i,lpha_j\}) = \mathbb{E}_{f_i, f_j}[\max(f_i - f^\star, f_j - f^\star, 0)]$ 

 $ig( egin{split} f_i \ f_j \end{pmatrix} \sim \mathcal{N}ig( ig( egin{split} \mu_i \ \mu_j \end{pmatrix}, \ ig( egin{split} \Sigma_{i,i} \ \Sigma_{i,j} \ \Sigma_{i,i} \end{pmatrix} ig)$ 



Using GP posteriors and utility functions

• 
$$lpha_{ ext{EI}}(lpha) = \mathbb{E}_f[\max(f-f^\star,0)]$$

• 
$$lpha_{ ext{EI}}(\{lpha_i,lpha_j\}) = \mathbb{E}_{f_i,\,f_j}[\max(f_i-f^\star,f_j-f^\star,0)]$$

• 
$$\alpha_{\mathrm{EI}}(\{st_1,\ldots,st_B\})=???$$





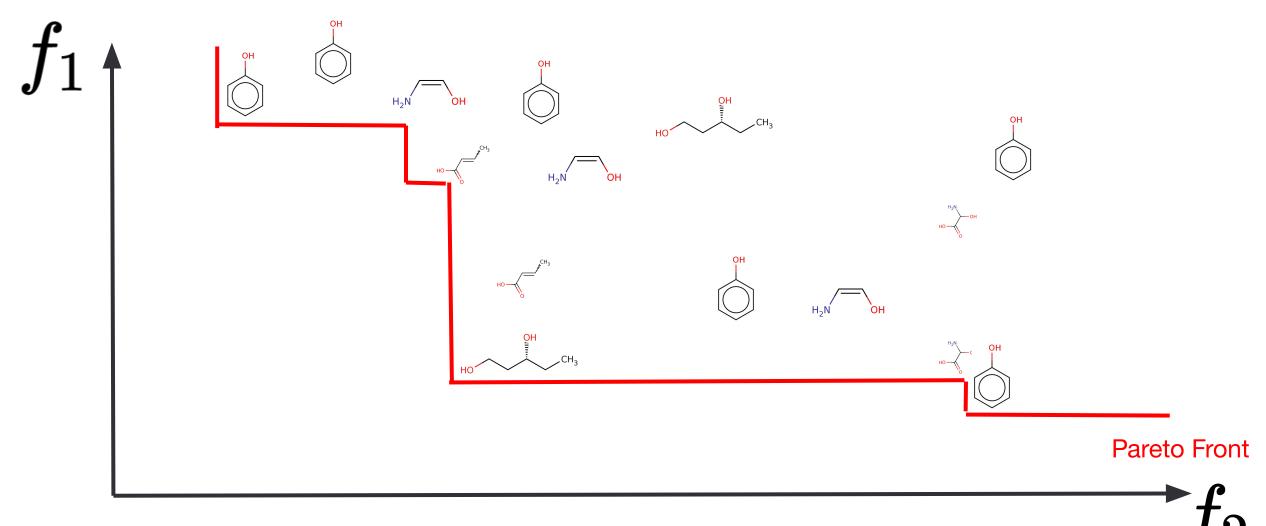
# Back to molecular design Multiple objectives



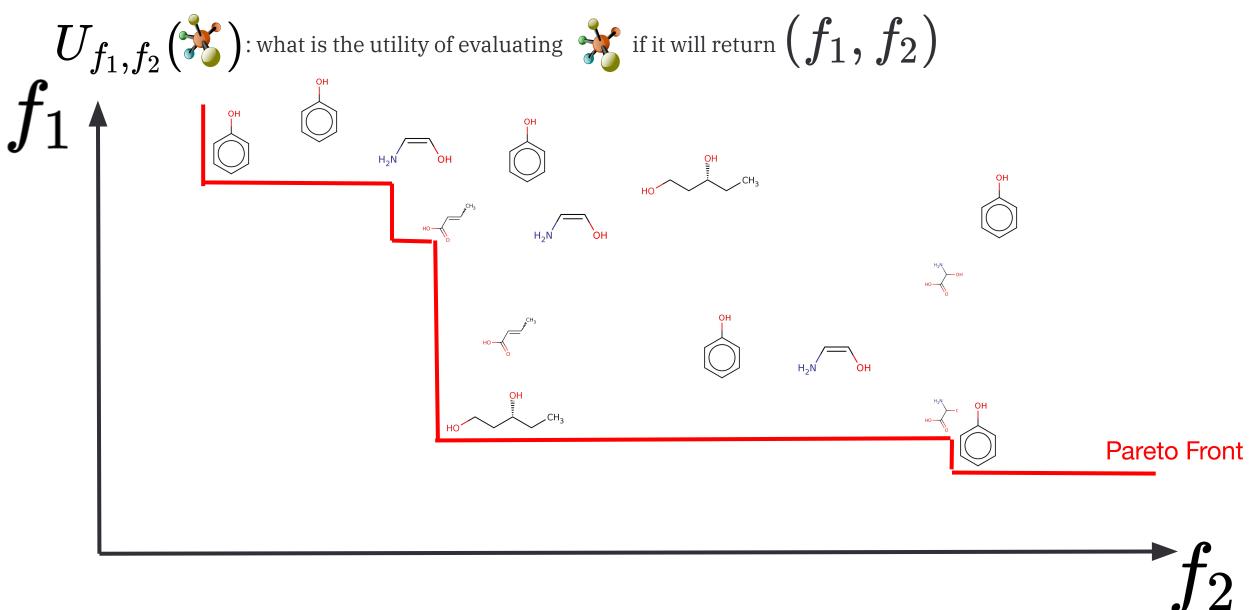
| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | • |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | • |
| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | • |

### Multi-objective Optimisation

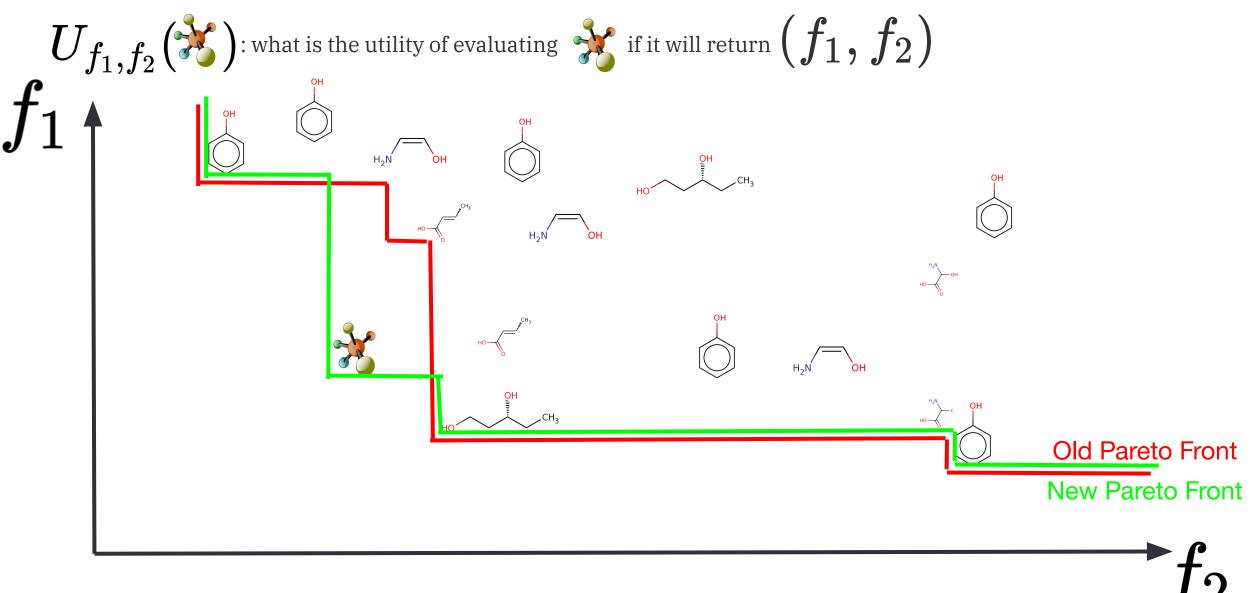
#### >1 competing objectives

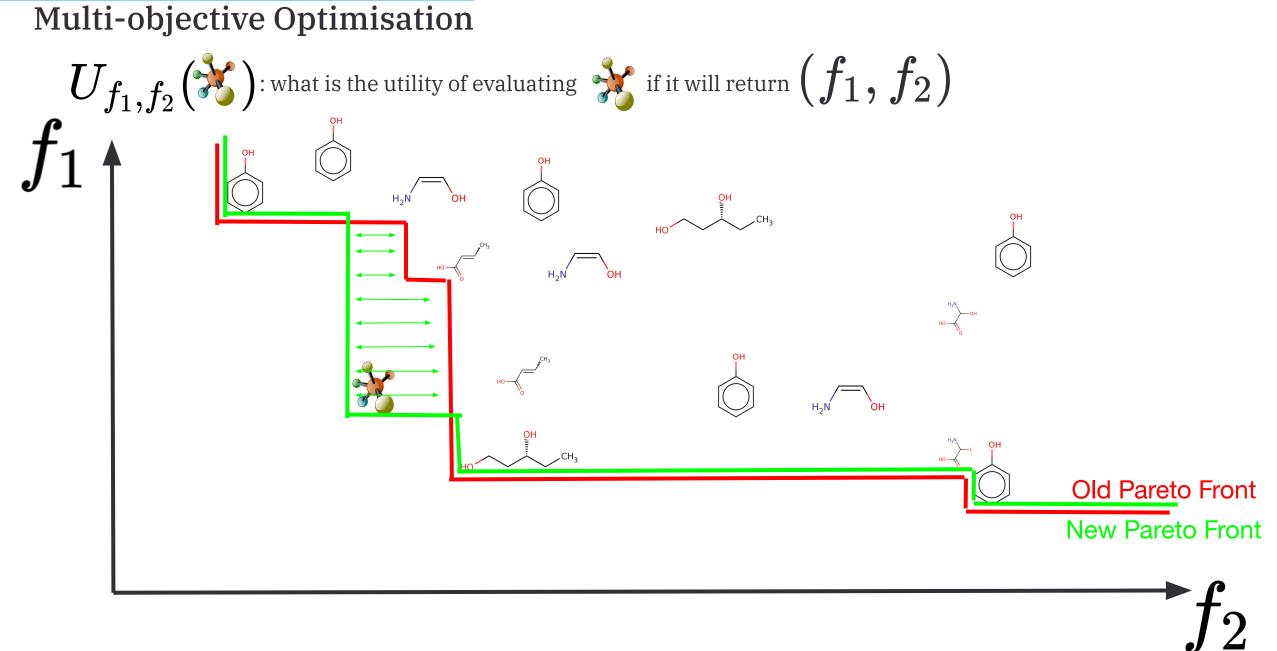


#### Multi-objective Optimisation

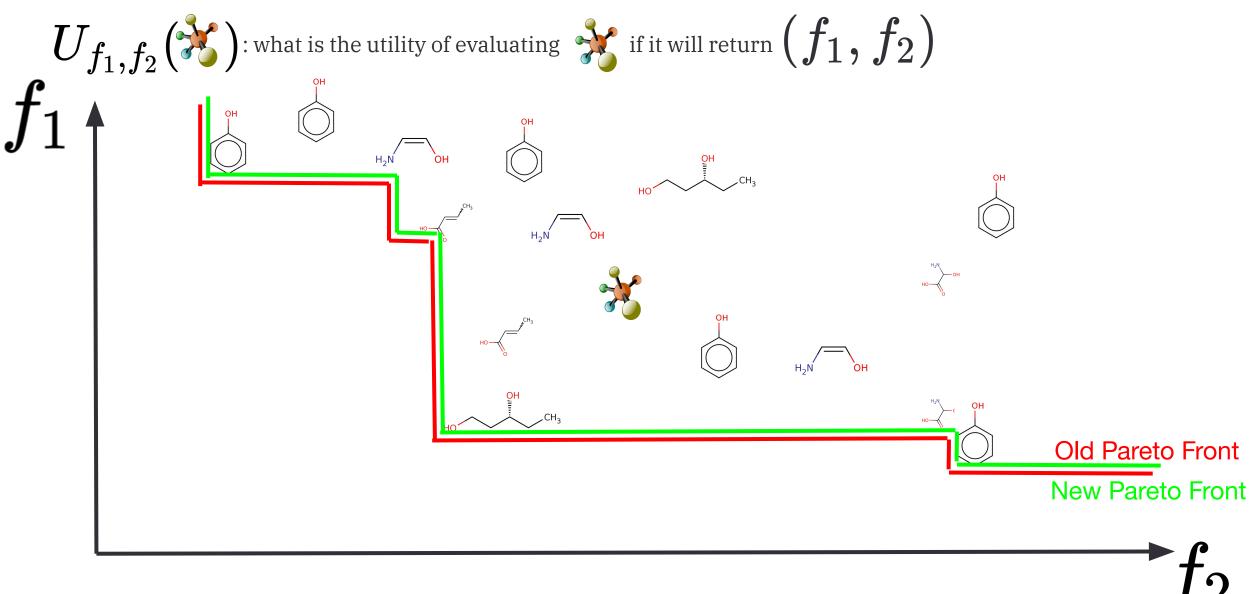


### Multi-objective Optimisation

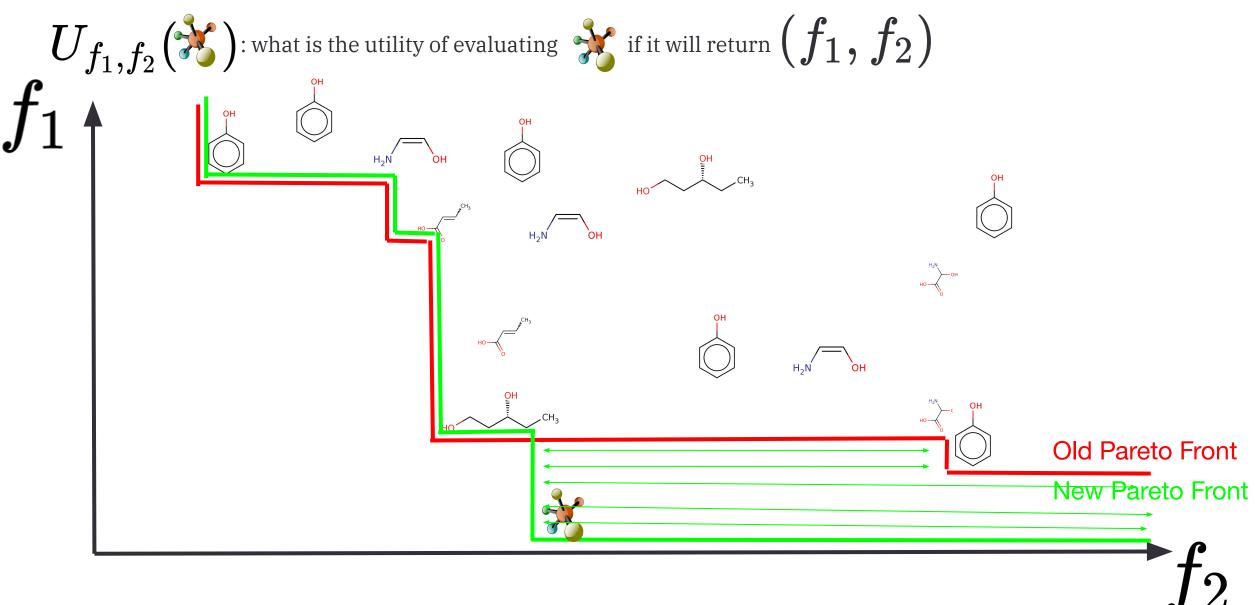




### **Multi-objective Optimisation**



### **Multi-objective Optimisation**



#### Multi-objective Optimisation

$$U_{f_1,f_2}(lpha)$$
 : what is the utility of evaluating  $\,\,lpha\,$  if it will return  $(f_1,f_2)$ 

- Use expected hyper-volume improvement  $~~lpha_{ ext{EHVI}}(lpha) = \mathbb{E}_{f_1,f_2}(U_{f_1,f_2}(lpha))$ 

$$f_1 \sim \mathcal{N}ig(\mu_1,\,\sigma_1^2ig) \ f_2 \sim \mathcal{N}ig(\mu_2,\,\sigma_2^2ig)$$

#### Multi-objective Optimisation

$$U_{f_1,f_2}(lpha)$$
 : what is the utility of evaluating  $\,\,lpha\,$  if it will return  $(f_1,f_2)$ 

- Use expected hyper-volume improvement  $~~lpha_{ ext{EHVI}}(lpha) = \mathbb{E}_{f_1,f_2}(U_{f_1,f_2}(lpha))$ 

$$f_1 \sim \mathcal{N}ig(\mu_1,\,\sigma_1^2ig) \ f_2 \sim \mathcal{N}ig(\mu_2,\,\sigma_2^2ig)$$

$$\alpha_{\mathrm{EHVI}}(\{ i, i, j \} \} = ???$$





A more sophisticated acquisition function?

Entropy Search





**Quick Recap** 

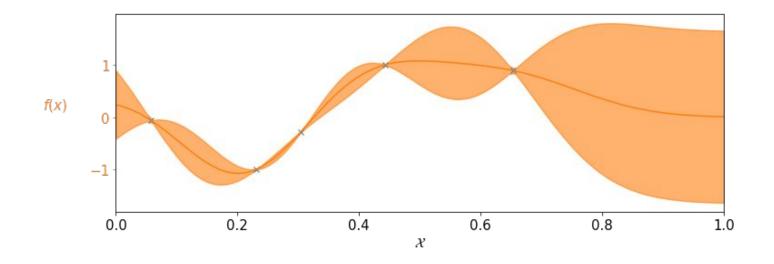
 $\mathbf{x}^* = \arg \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x})$ 



**Quick Recap** 

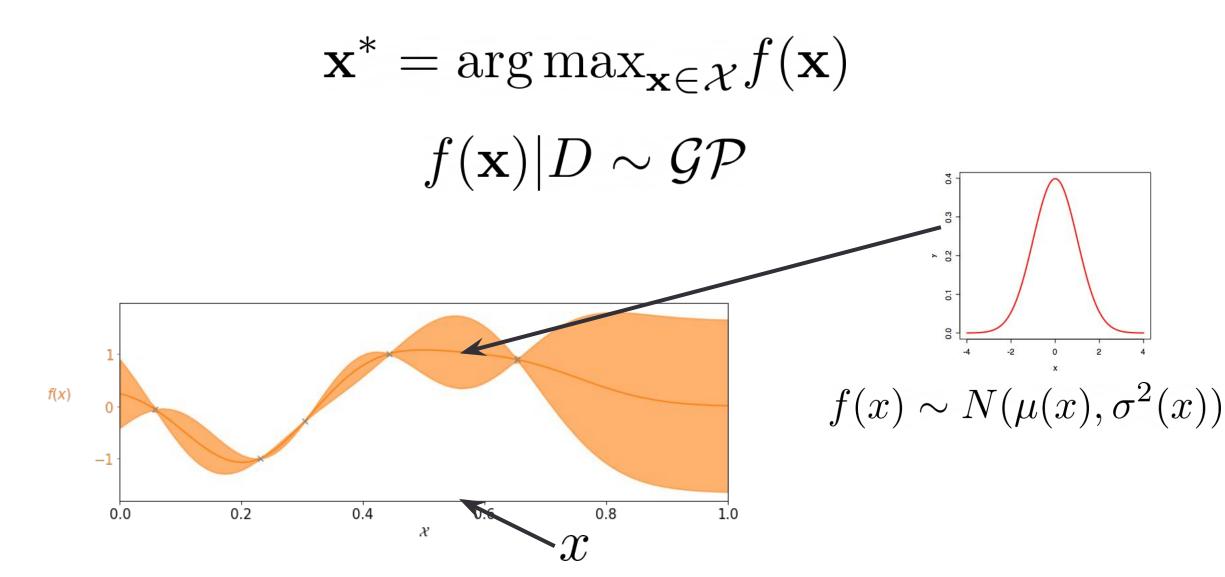
$$\mathbf{x}^* = \arg \max_{\mathbf{x} \in \mathcal{X}} f(\mathbf{x})$$

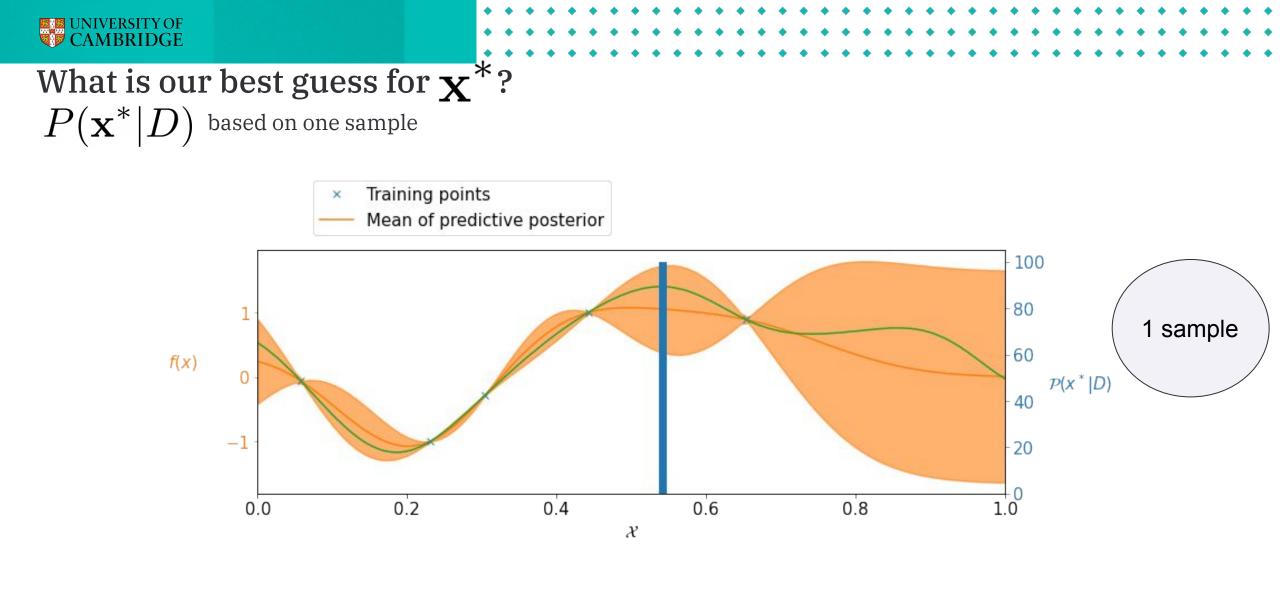
 $f(\mathbf{x})|D \sim \mathcal{GP}$ 

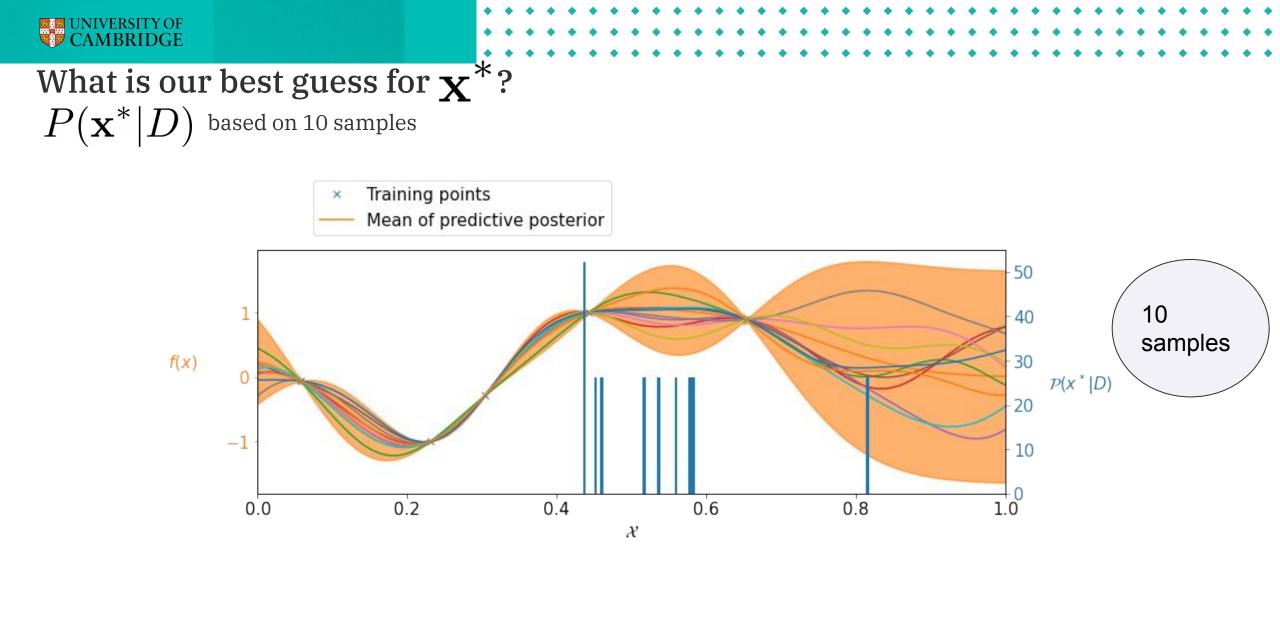


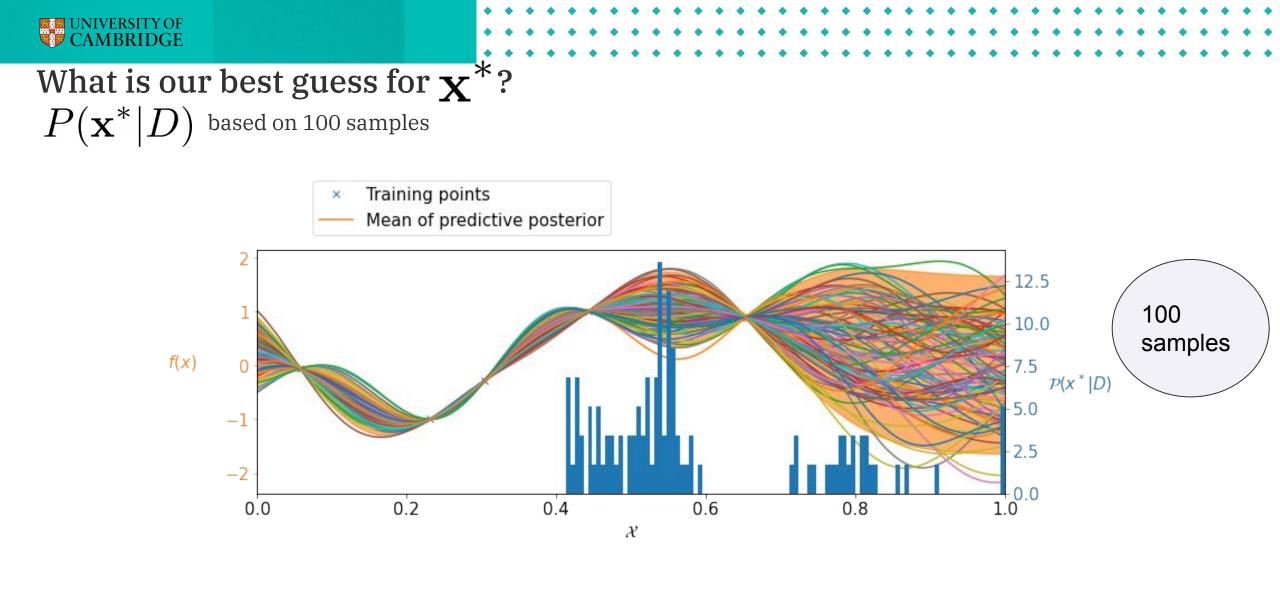


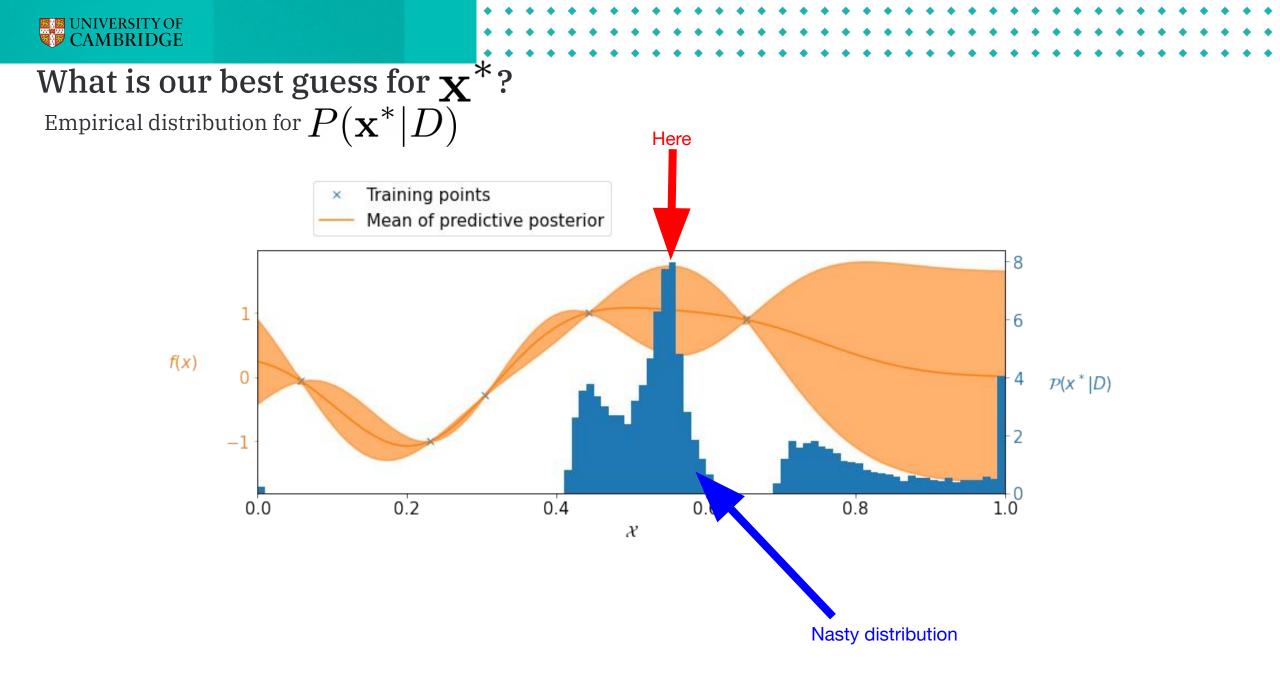
**Quick Recap** 











## Where shall we evaluate next?

We want to learn about  $\mathbf{X}^{*}$ 

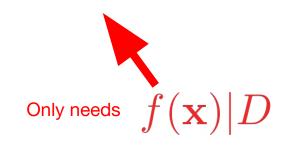
• Expected Improvement (EI) maximises  $\alpha_{EI}(\mathbf{x}) = E[\max(f(\mathbf{x}) - f^*, 0)]$ 



## Where shall we evaluate next?

We want to learn about  $\mathbf{X}^{*}$ 

• Expected Improvement (EI) maximises  $\alpha_{EI}(\mathbf{x}) = E[\max(f(\mathbf{x}) - f^*, 0)]$ 

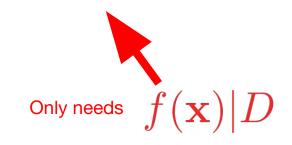




## Where shall we evaluate next?

We want to learn about  $\mathbf{X}^{*}$ 

• Expected Improvement (EI) maximises  $\alpha_{EI}(\mathbf{x}) = E[\max(f(\mathbf{x}) - f^*, 0)]$ 



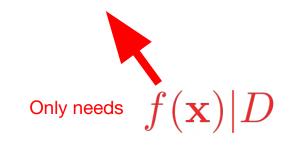
# Does not use full knowledge of $P(\mathbf{x}^*|D)$



## Where shall we evaluate next?

We want to learn about  $\mathbf{X}^{*}$ 

• Expected Improvement (EI) maximises  $\alpha_{EI}(\mathbf{x}) = E[\max(f(\mathbf{x}) - f^*, 0)]$ 



# Does not use full knowledge of $P(\mathbf{x}^*|D)$

**Entropy search** seeks to reduce our uncertainty in  $P(\mathbf{x}^*|D)$ 



| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | • |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | • |
| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | • |





$$\operatorname{Var}(X) = E\left[(X - \mu)^2\right]$$





$$\operatorname{Var}(X) = E\left[(X - \mu)^2\right]$$

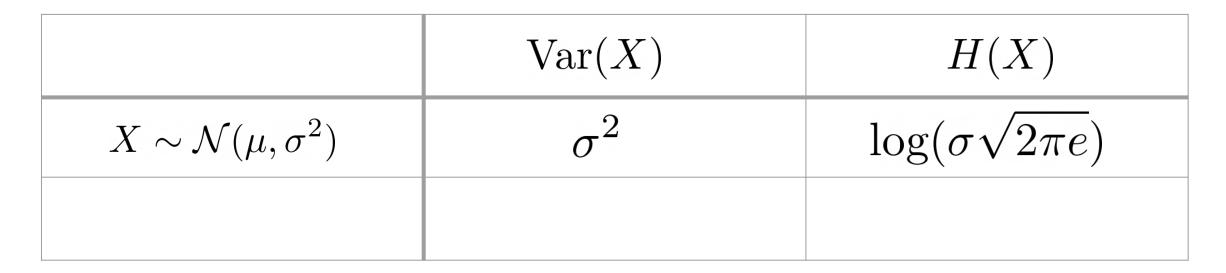
$$H(X) = E\left[-\log(p(X))\right]$$





$$\operatorname{Var}(X) \stackrel{\scriptscriptstyle{\mathrm{r}}}{=} E\left[(X-\mu)^2\right]$$

$$H(X) = E\left[-\log(p(X))\right]$$

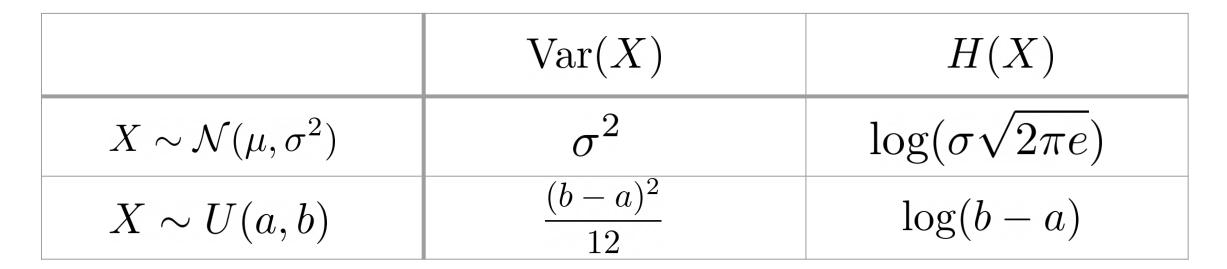






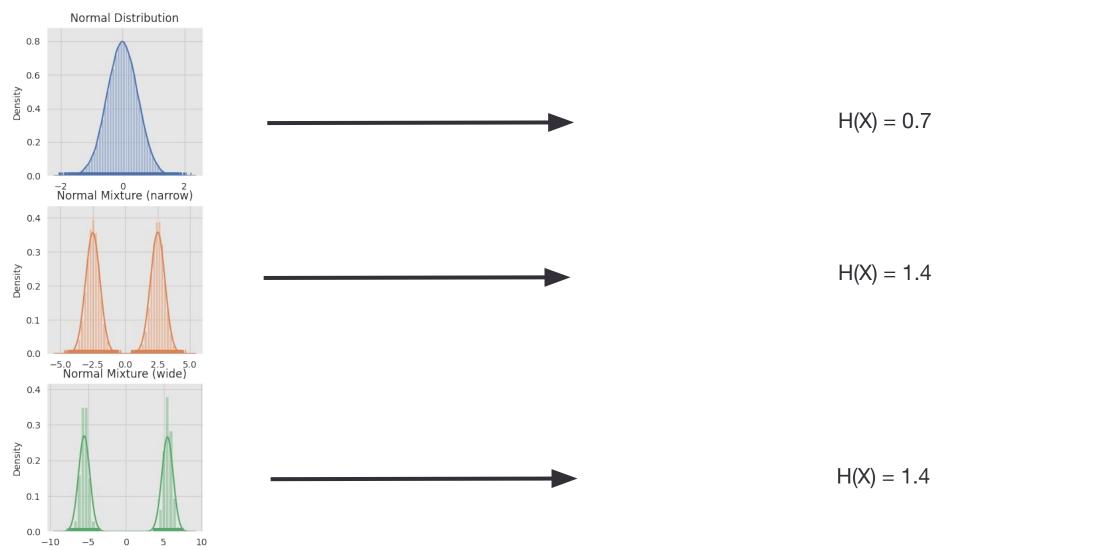
$$\operatorname{Var}(X) \stackrel{\scriptscriptstyle{\mathrm{r}}}{=} E\left[(X-\mu)^2\right]$$

$$H(X) = E\left[-\log(p(X))\right]$$



## How to measure uncertainty?

### Should we use entropy?

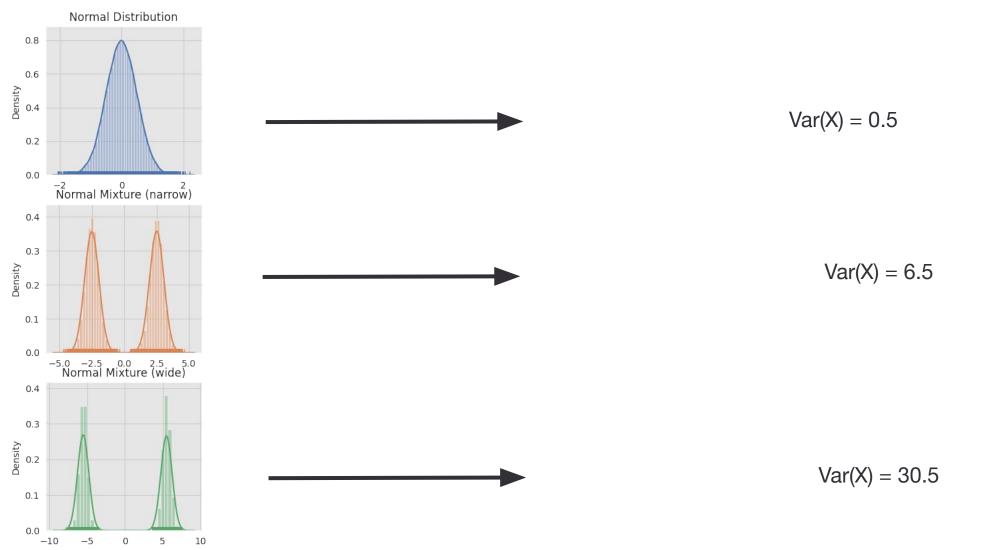


 $H(X) = E\left[-\log(p(X))\right]$ 

# $\operatorname{Var}(X) = E\left[(X - \mu)^2\right]$

# How to measure uncertainty?

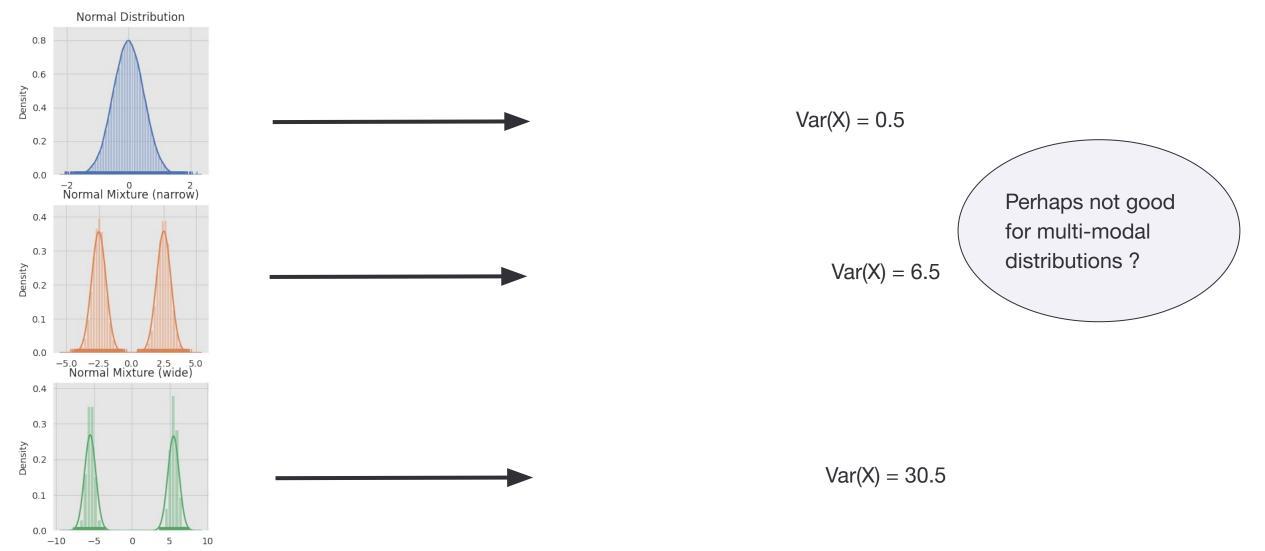
## Should we use variance (i.e. dispersion)?



# $\operatorname{Var}(X) = E\left[(X - \mu)^2\right]$

# How to measure uncertainty?

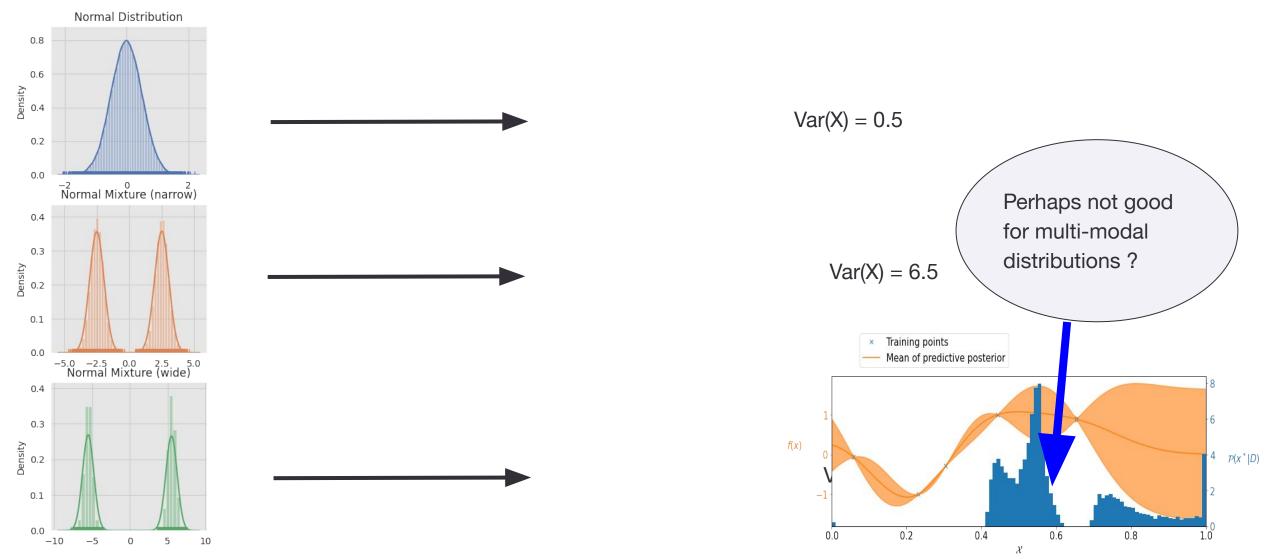
## Should we use variance (i.e. dispersion)?



# $\operatorname{Var}(X) = E\left[(X - \mu)^2\right]$

# How to measure uncertainty?

## Should we use variance (i.e. dispersion)?





| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ |
| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ |

# Entropy Search Reduce global uncertainty in $P(\mathbf{x}^*|D)$

# Entropy Search Reduce global uncertainty in $P(\mathbf{x}^*|D)$

## How?

• Measure uncertainty by differential entropy  $H(\mathbf{x}^*|D) = -E_{\mathbf{x} \sim \mathbf{x}^*|D}[\log(p(\mathbf{x}))]$ 

# Entropy Search Reduce global uncertainty in $P(\mathbf{x}^*|D)$

## How?

- Measure uncertainty by differential entropy  $H(\mathbf{x}^*|D) = -E_{\mathbf{x} \sim \mathbf{x}^*|D}[\log(p(\mathbf{x}))]$
- Make evaluation that provides the largest expected reduction in entropy

$$\alpha_{ES}(\mathbf{x}) = H(\mathbf{x}^*|D) - E_y[H(\mathbf{x}^*|D \cup \{y, \mathbf{x}\})]$$

# Entropy Search Reduce global uncertainty in $P(\mathbf{x}^*|D)$

## How?

- Measure uncertainty by differential entropy  $H(\mathbf{x}^*|D) = -E_{\mathbf{x} \sim \mathbf{x}^*|D}[\log(p(\mathbf{x}))]$
- Make evaluation that provides the largest expected reduction in entropy

$$\alpha_{ES}(\mathbf{x}) = H(\mathbf{x}^* | D) - E_y[H(\mathbf{x}^* | D \cup \{y, \mathbf{x}\})]$$
Current uncertainty
Current uncertainty
Expected uncertainty after collecting evaluation  $\mathcal{Y}$  at location  $\mathbf{X}$ 

# Entropy Search Reduce global uncertainty in $P(\mathbf{x}^*|D)$

## How?

- Measure uncertainty by differential entropy  $H(\mathbf{x}^*|D) = -E_{\mathbf{x} \sim \mathbf{x}^*|D}[\log(p(\mathbf{x}))]$
- Make evaluation that provides the largest expected reduction in entropy

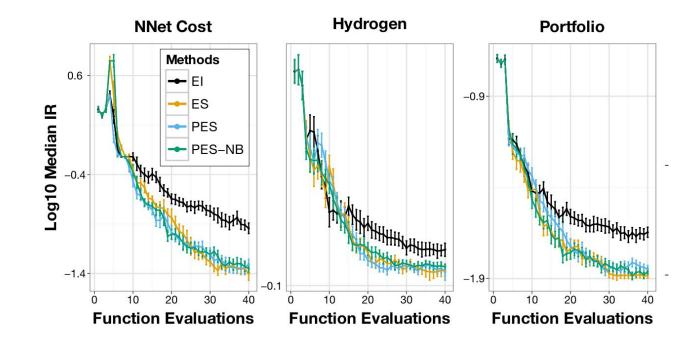
$$\alpha_{ES}(\mathbf{x}) = H(\mathbf{x}^*|D) - E_y[H(\mathbf{x}^*|D \cup \{y, \mathbf{x}\})]$$
Current uncertainty
Current uncertainty
Expected uncertainty after collecting evaluation  $\mathcal{Y}$  at location  $\mathbf{X}$ 

Fiendishly difficult to calculate!

• What is  $H(\mathbf{x}^*|D)$ ? • What is  $H(\mathbf{x}^*|D, \{y, \mathbf{x}\})$ ???

# It can be worth calculating these horrible quantities

They can provide highly efficient optimization



For details see

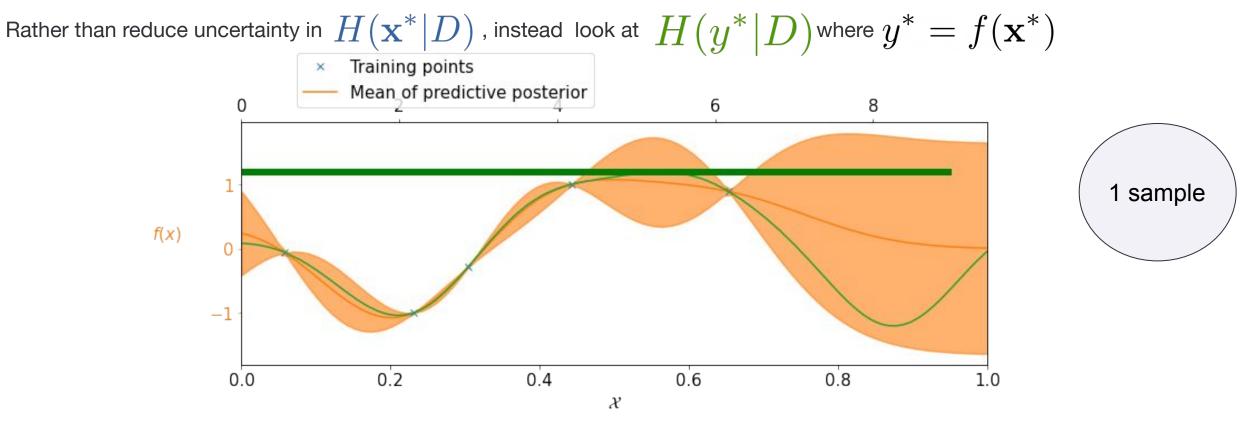
- Entropy Search is  $O(n^2 e^{2d} + e^{3d})$ (Henning and Schuler, 2012) Predictive Entropy Search is  $O(n^2 e^{2d} + n^3 e^d)$  (Hernandez-Lobato et al. 2014)



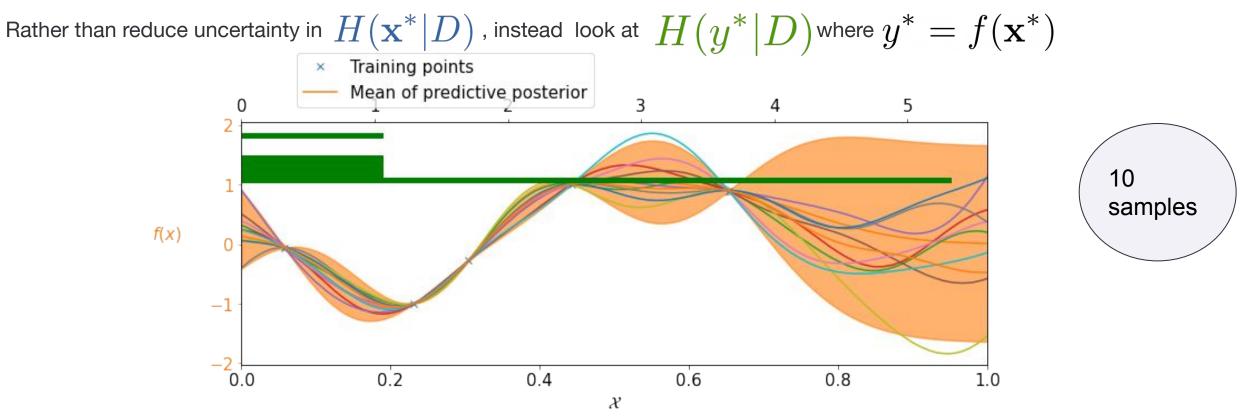
Min-value Entropy Search

Rather than reduce uncertainty in  $H(\mathbf{x}^*|D)$  , instead look at  $H(y^*|D)$  where  $y^*=f(\mathbf{x}^*)$ 

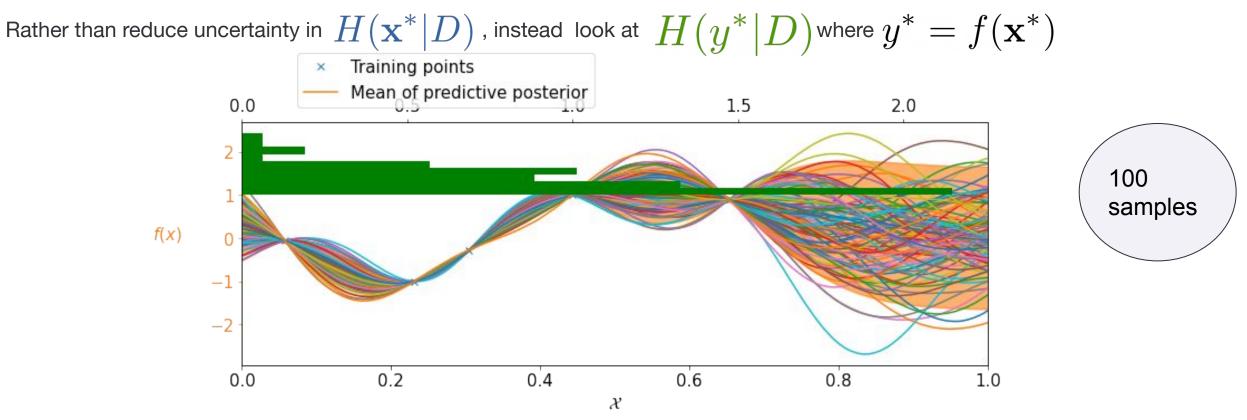




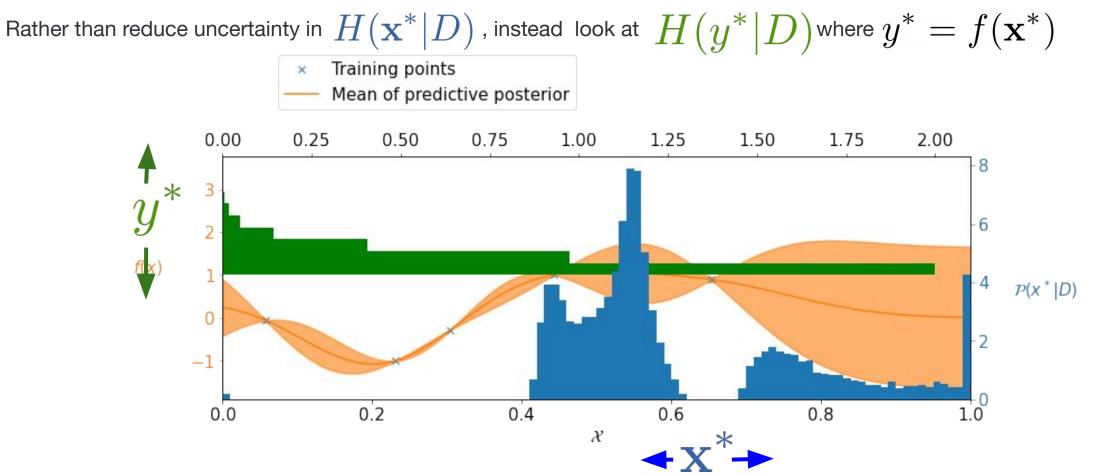






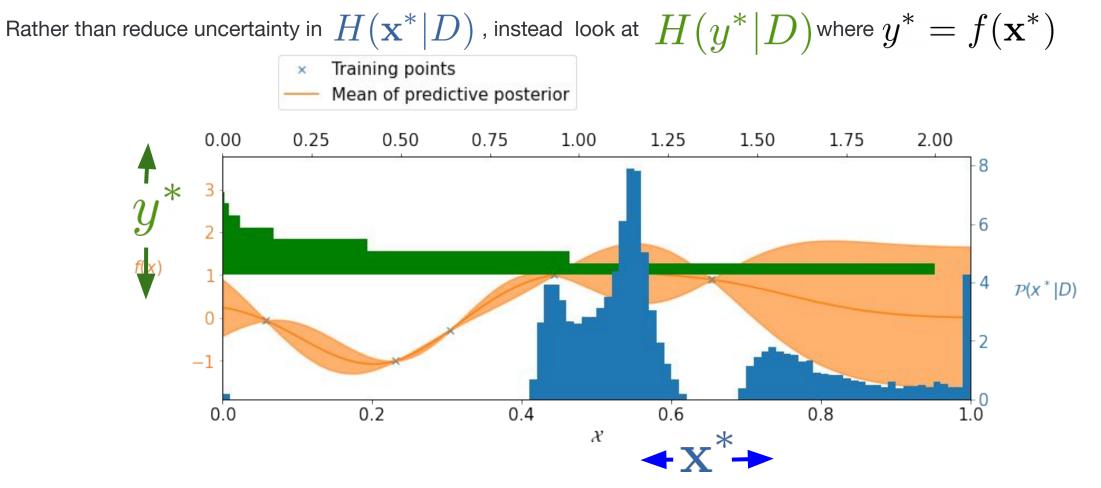


# There is a better way!



# There is a better way!

Min-value Entropy Search



 $\alpha_{MES}(\mathbf{x}) = H(y|D) - E_{y^*|D}[y|D \cup y^*]$ 

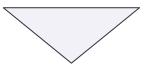




Min-value Entropy Search

Rather than reduce uncertainty in  $H(\mathbf{x}^*|D)$  , instead look at  $H(y^*|D)$  where  $y^*=f(\mathbf{x}^*)$ 

$$\alpha_{\text{MES}}(\mathbf{x}) = H(y^*|D) - E_{y|D} \left[ H(y^*|D\bigcup(y,\mathbf{x})) \right]$$





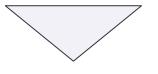
Current uncertainty

Expected uncertainty after the evaluation



Rather than reduce uncertainty in 
$$H(\mathbf{x}^*|D)$$
 , instead look at  $\,H(y^*|D)$  where  $y^*=f(\mathbf{x}^*)$ 

$$\alpha_{\text{MES}}(\mathbf{x}) = H(y^*|D) - E_{y|D} \left[ H(y^*|D\bigcup(y,\mathbf{x})) \right]$$





Current uncertainty

Expected uncertainty after the evaluation

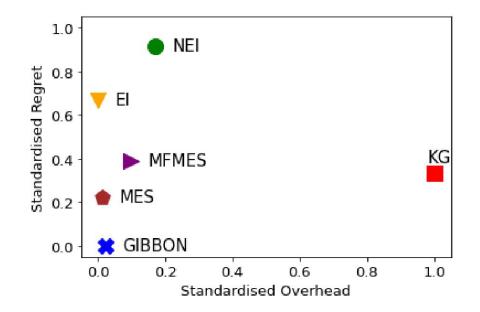
Crucially 
$$\mathbf{y}^* \in R$$
 , whereas  $\mathbf{x}^* \in R^d$ 



| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ |
| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | • |

# MES in practice

Highly effective optimization at low cost!

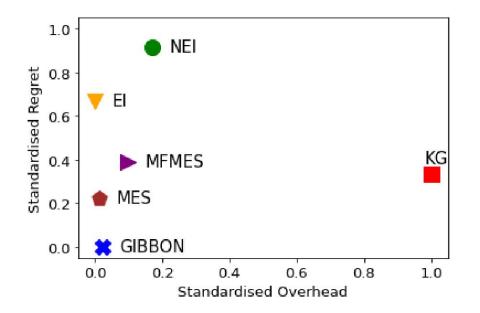




| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | • |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ |
| ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ | ٠ |

# **MES in practice**

Highly effective optimization at low cost!



- Max-Value Entropy Search is  $O(n^2 e^d)$  for noiseless optimisation (Wang and Jegelka, 2017).
- MUMBO is  $O(n^2 e^d)$  for noisy optimisation (Moss et al., 2020) GIBBON is  $O((n^2 + B^2)e^d + B^3)$  for batches of size B (Moss et al. 2021)

# Thanks for listening

