How to Optimize Anything

Tim Head

21 July 2017

How to Make The Best

Tim Head
tim@wildtreetech.com
@betatim

Beer recipes

Your task: brew the most tasty beer possible.

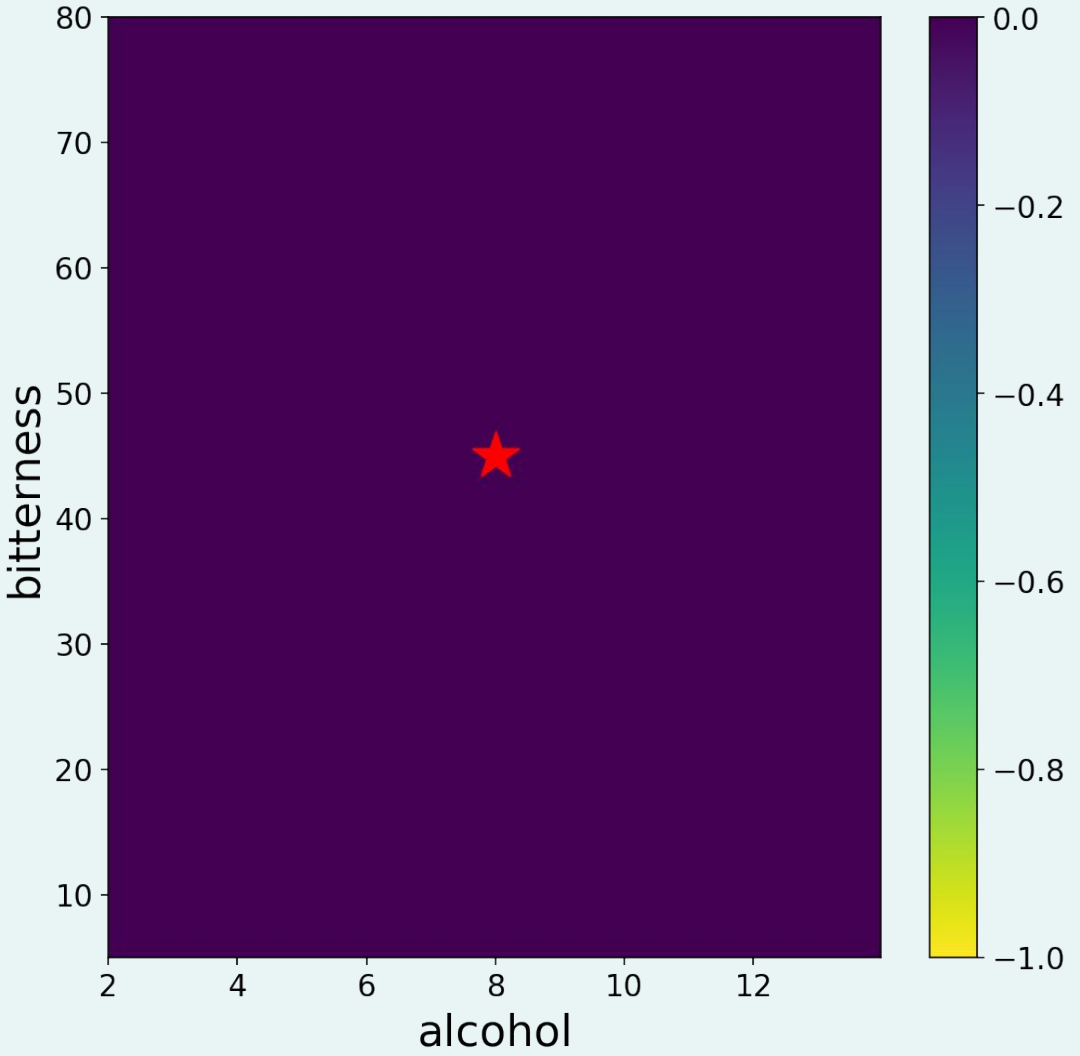
Many parameters you can tweak, for simplicity we let’s pretend there are only two: **alcohol content** and **bitterness**.

How do you find the best combination?

Beer parameters

At the start we don’t know anything about how the parameters really influence the score.

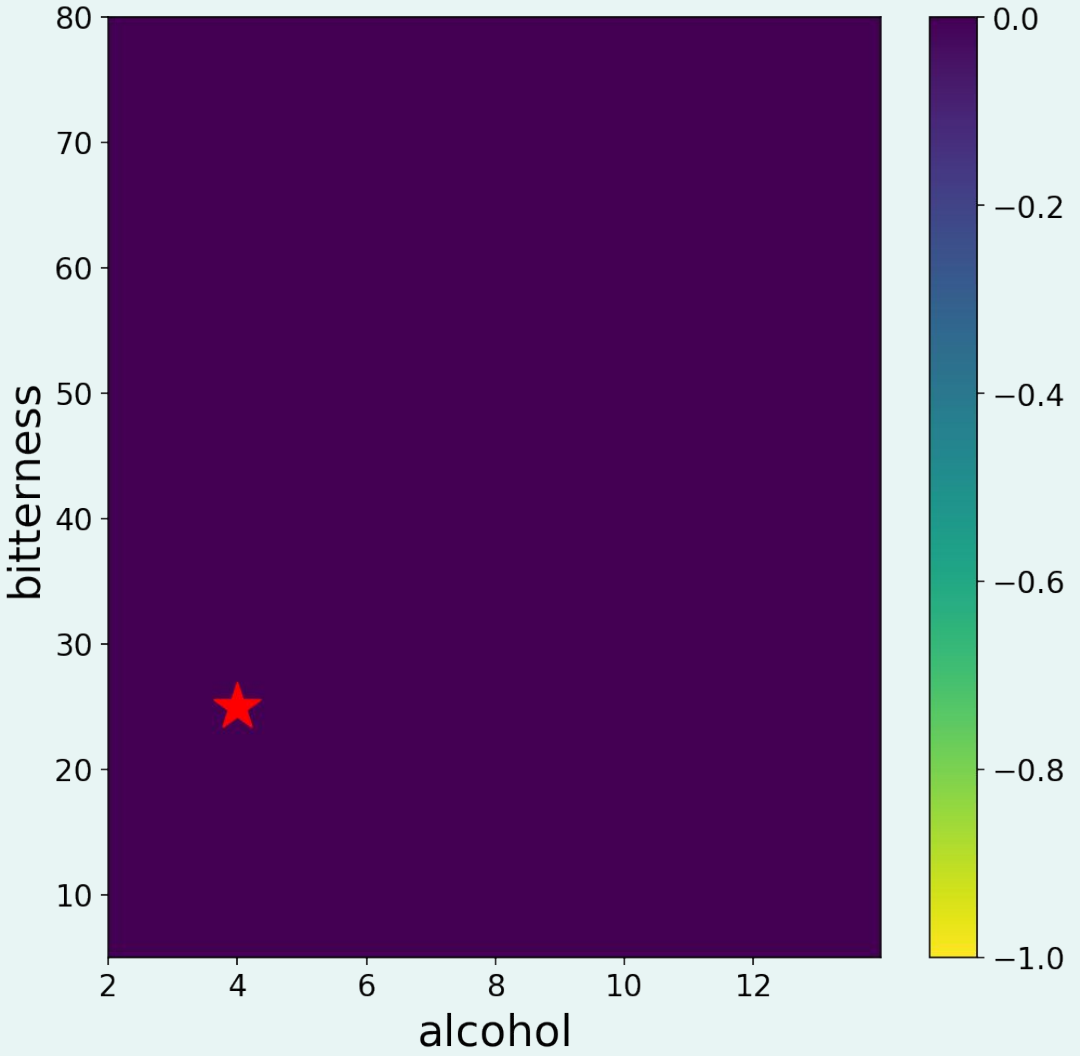
The task is to find the best score in as few trials as possible.



Beer parameters

At the start we don’t know anything about how the parameters really influence the score.

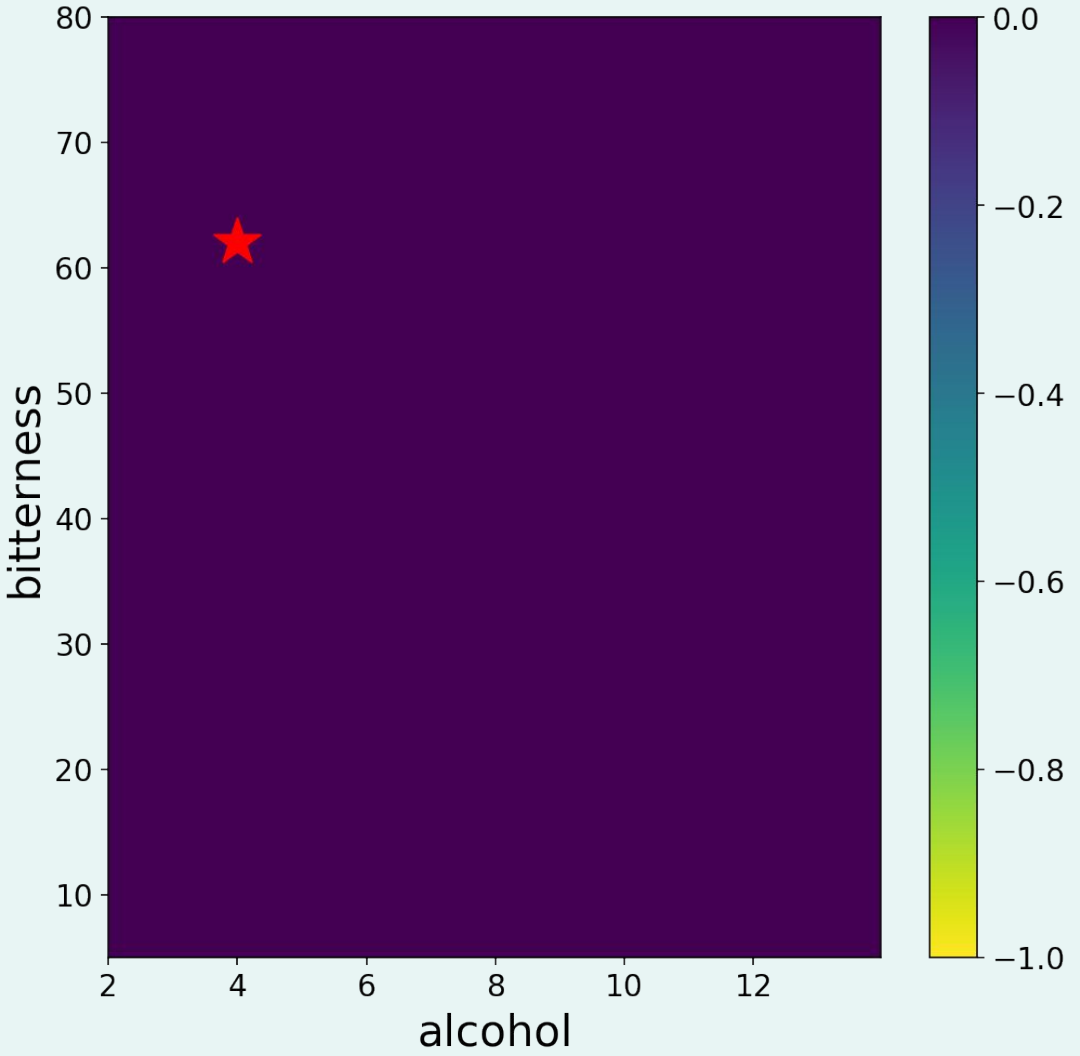
The task is to find the best score in as few trials as possible.



Beer parameters

At the start we don’t know anything about how the parameters really influence the score.

The task is to find the best score in as few trials as possible.



Evaluating a recipe is expensive

How to score a beer:

- Buy ingredients
- Brew it
- ...wait...
- Find expert panel and collect scores

This means we can't try a large number of combinations, have to be smart.

This is an optimization problem, but what is the *objective function*?

Brewery-in-a-laptop

I have a simulation of the brewing and tasting process on my laptop.

For a given set of **alcohol content** and **bitterness** it returns a score.

```
beer_score(alcohol, bitterness)
-> score
```

Demo

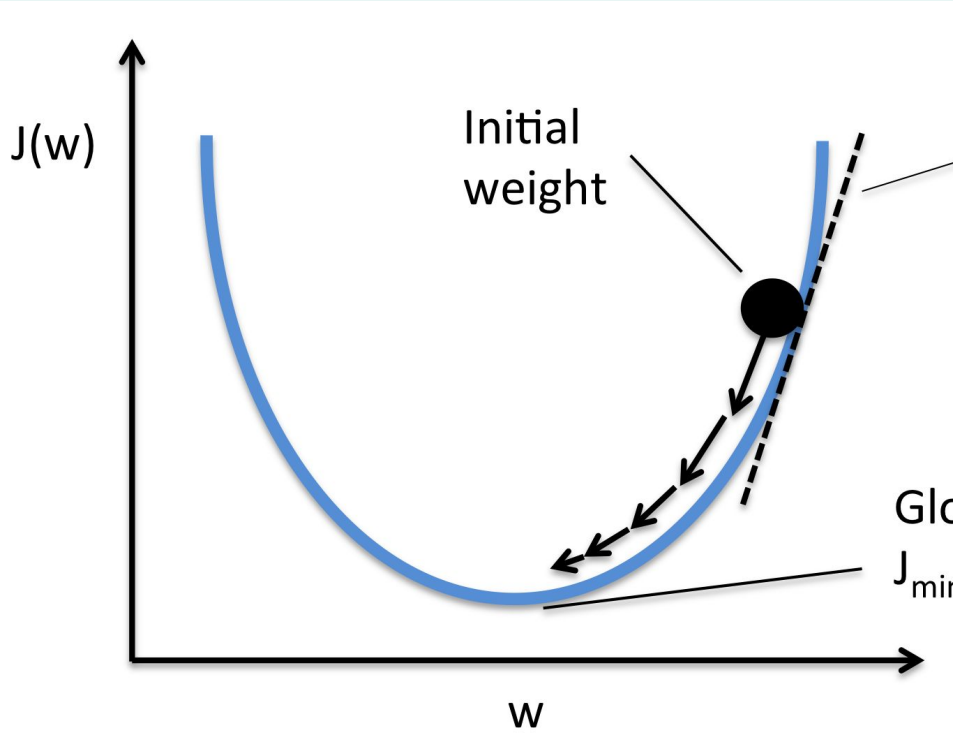
Gradient descent

Gradient descent algorithms need to compute the gradient.

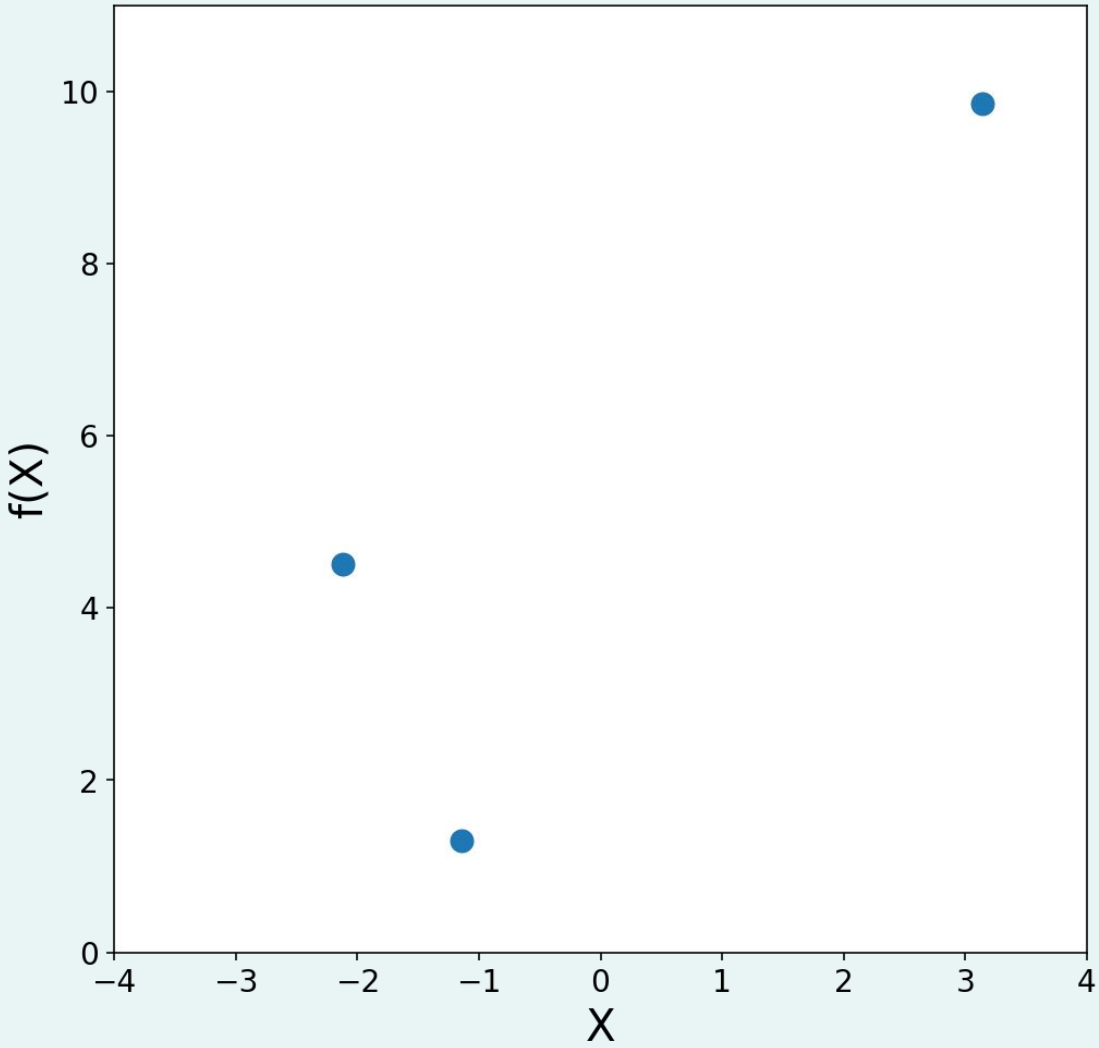
Computing the gradient numerically is expensive.

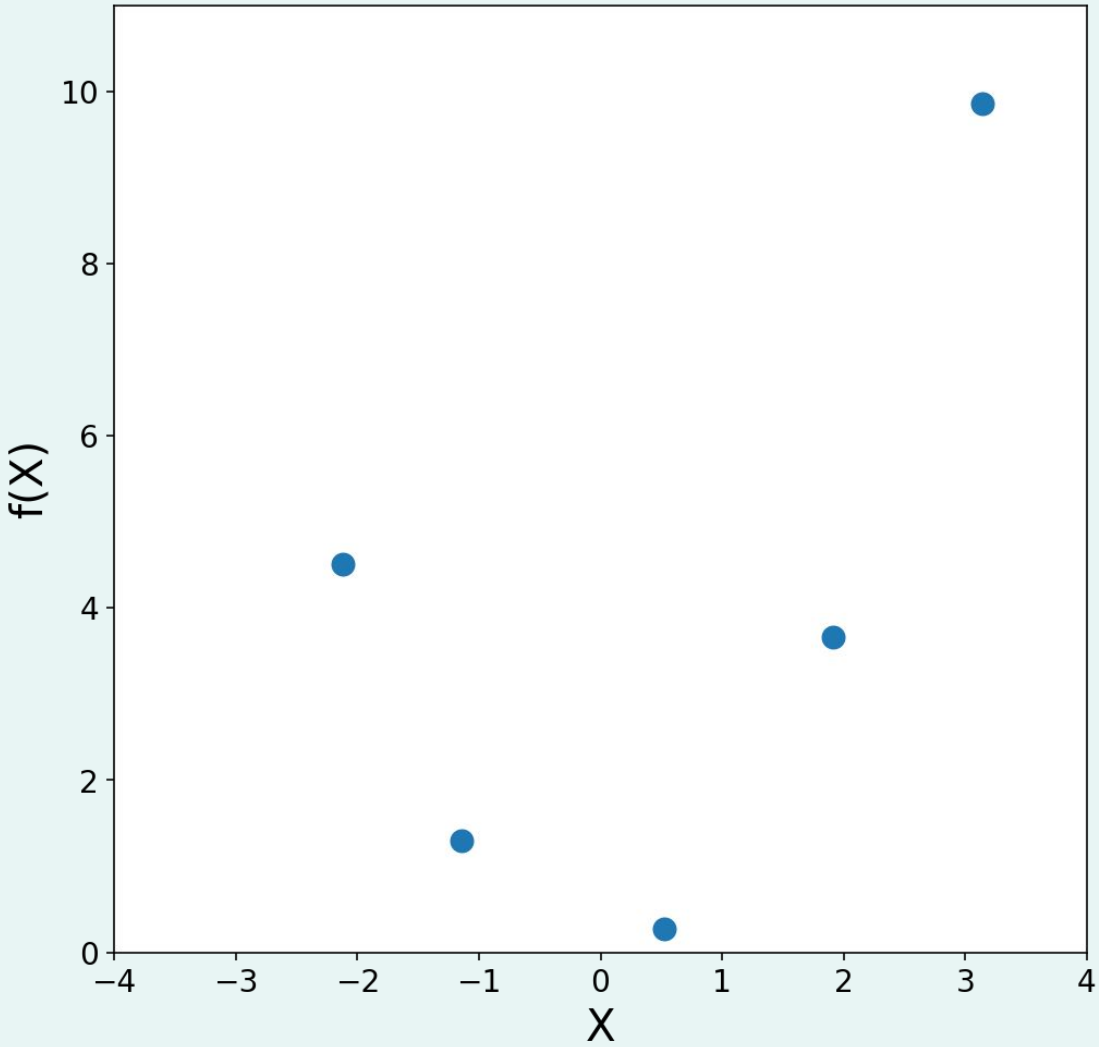
We waste a lot of evaluations on this:

$$\frac{\partial f}{\partial x} \approx \frac{f(a+\Delta) - f(a)}{\Delta}$$



Bayesian Optimisation


10
8
6
4
2
0
f(X)
−4
−3
−2
−1
0
1
2
3
4
X


10
8
6
4
2
0
f(X)
−4
−3
−2
−1
0
1
2
3
4
X

The Big Idea

Use the points for which we have evaluated f to predict where the minimum is.

Evaluate f at that point.

Update our model to make a new prediction.

$$x^* = \arg\max_x f(x)$$

- f is a black box function, with no closed form nor gradients.
- f is expensive to evaluate.
- You may only have noisy observations of f.

If you do not have these constraints, *do not* use Bayesian optimization.

Bayesian optimisation loop

Given a set of observations: $\{(x_i, f(x_i))|i = 1, \ldots, t\}$

1. Fit a regression model to the observations
2. Minimize a cheap “acquisition function” ***u(x)*** based on the model to find the next point $x_{t+1} = \arg\max_x u(x)$
3. Evaluate ***f*** at that point
4. Repeat by going to step 1.

Demo

Ingredients needed

Given a set of observations $\{(x_i, f(x_i))|i = 1, \ldots, t\}$

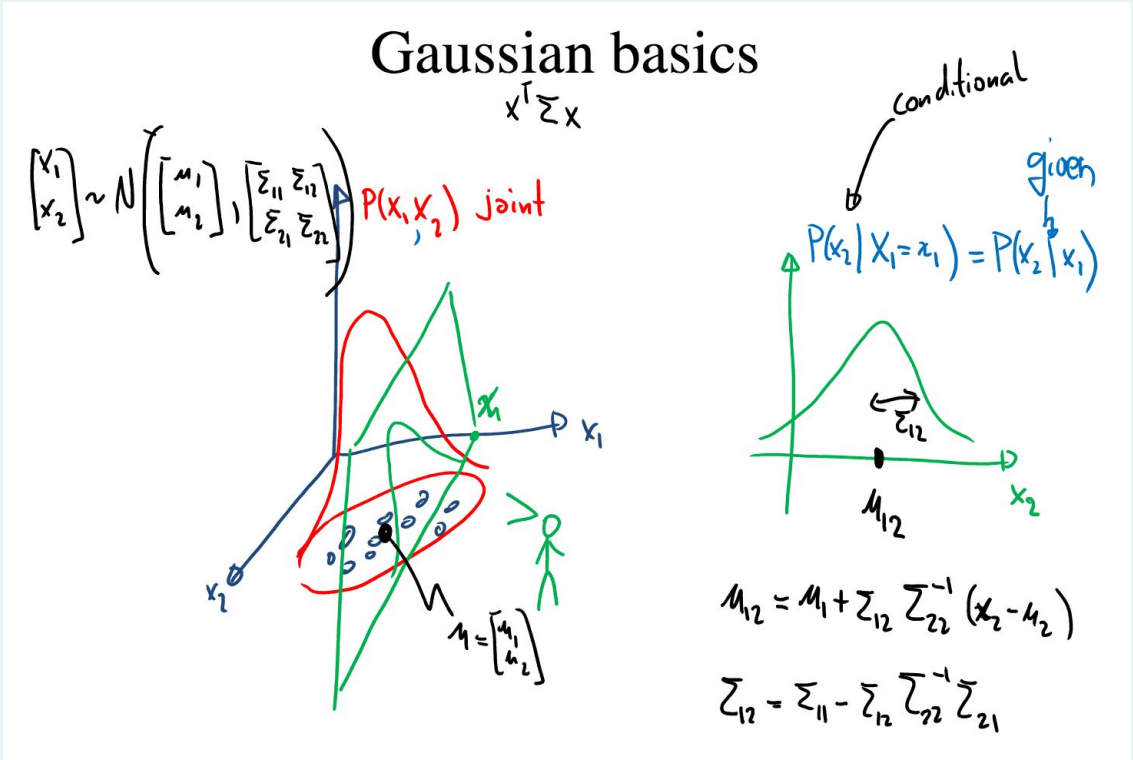
1. Fit a ***regression model*** to the observations
2. Minimize a cheap "***acquisition function***" ***u(x)*** based on the model to find the next point $x_{t+1} = \arg\max_x u(x)$
3. Evaluate ***f*** at that point
4. Repeat by going to step 1.

Acquisition functions

Demo

Regression with uncertainties

Detour: Gaussian Processes



If you want details watch Nando de Freitas - Intro to GPs it is excellent!

Multivariate Gaussian Theorem (see KPM)

Theorem 4.2.1 (Marginals and conditionals of an MVN). *Suppose* $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$ *is jointly Gaussian with parameters*

$$\boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{pmatrix}, \quad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}, \quad \boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \boldsymbol{\Lambda}_{11} & \boldsymbol{\Lambda}_{12} \\ \boldsymbol{\Lambda}_{21} & \boldsymbol{\Lambda}_{22} \end{pmatrix} \tag{4.12}$$

Then the marginals are given by

$$\begin{aligned} p(\mathbf{x}_1) &= \mathcal{N}(\mathbf{x}_1 | \boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}) \\ p(\mathbf{x}_2) &= \mathcal{N}(\mathbf{x}_2 | \boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}) \end{aligned}$$

and the posterior conditional is given by

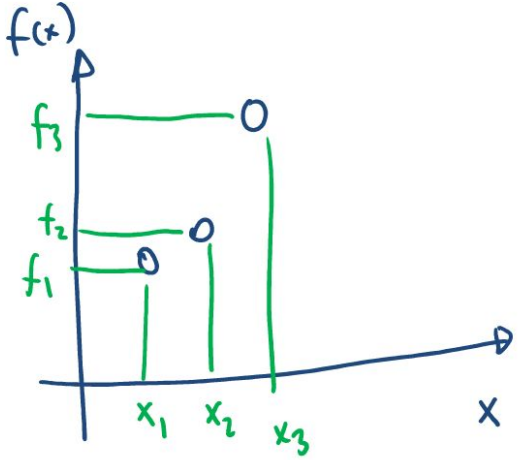
$$\begin{aligned} p(\mathbf{x}_1|\mathbf{x}_2) &= \mathcal{N}(\mathbf{x}_1 | \boldsymbol{\mu}_{1|2}, \boldsymbol{\Sigma}_{1|2}) \\ \boldsymbol{\mu}_{1|2} &= \boldsymbol{\mu}_1 + \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}(\mathbf{x}_2 - \boldsymbol{\mu}_2) \\ &= \boldsymbol{\mu}_1 - \boldsymbol{\Lambda}_{11}^{-1}\boldsymbol{\Lambda}_{12}(\mathbf{x}_2 - \boldsymbol{\mu}_2) \\ &= \boldsymbol{\Sigma}_{1|2}\left(\boldsymbol{\Lambda}_{11}\boldsymbol{\mu}_1 - \boldsymbol{\Lambda}_{12}(\mathbf{x}_2 - \boldsymbol{\mu}_2)\right) \\ \boldsymbol{\Sigma}_{1|2} &= \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} = \boldsymbol{\Lambda}_{11}^{-1} \end{aligned}$$

If you want details watch Nando de Freitas - Intro to GPs it is excellent!

Gaussian basics

x's given

want to model f's



$$\begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} K_{11} & K_{12} & K_{13} \\ K_{21} & K_{22} & K_{23} \\ K_{31} & K_{32} & K_{33} \end{bmatrix} \right)$$

$$\sim \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 0.7 & 0.2 \\ 0.7 & 1 & 0.6 \\ 0.2 & 0.6 & 1 \end{bmatrix} \right)$$

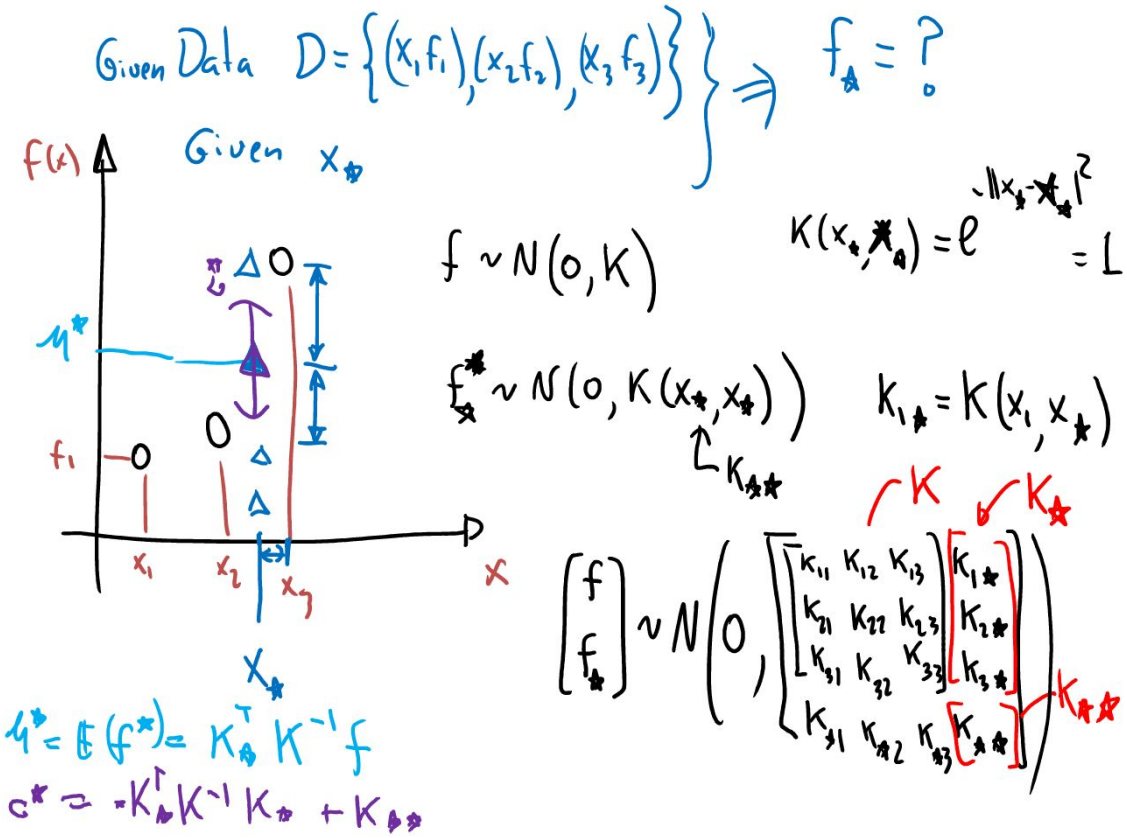
K

$$K_{ij} = e^{-\lambda \|x_i - x_j\|^2} = \begin{cases} 0 & \|x_i - x_j\| \to \infty \\ 1 & x_i = x_j \end{cases}$$

$$\underline{f} \sim N(0, K)$$

If you want details watch Nando de Freitas - Intro to GPs it is excellent!

Gaussian basics



If you want details watch Nando de Freitas - Intro to GPs it is excellent!

Demo

Back to the brewery

Demo

Applications beyond beer

Any “algorithm” that can be “configured”

What is an algorithm?

- a random forest,
- your analysis,
  - Which channels and cuts give overall smallest uncertainty?
- layout of a detector,
  - Where to place each sub-detector, what resolution does each one need?
- a beer recipe,
- mSSM parameter space,
  - where is the best fit point? Which point to exclude next for most bang-for-buck?

Material

wildtreetech/bayesian-optimisation

??!?

Tim Head
tim@wildtreetech.com
@betatim

Tim Head
tim@wildtreetech.com
@betatim