Introduction to Gaussian Processes (GPs):

Motivation, Basics, and Applications

Miles Woollacott
miles.woollacott@gmail.com

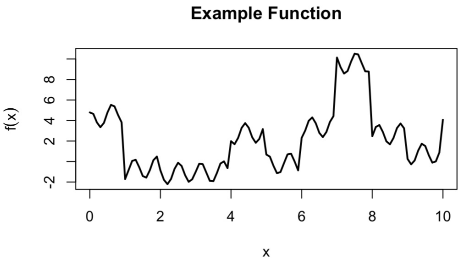
Summer 2025

Outline

Outline

Example Function

Suppose we want to find the maximum value of this function, $f(x)$ for $0 < x < 10$:



This is easy, because we have a graph of the function, and we can see that the maximum happens right around when $x = 7.6$.

If only it was that easy. . .

- How did we obtain a plot of that earlier function? We obtain the graph of $f(x)$ by sampling in a dense grid for $0 < x < 10$.
  - $f(x)$ doesn't take too long to run for a single point, so using a dense grid doesn't take that long, either.
  - What do we do if it takes a whole day to obtain a value of $f(x)$ for a *single* x?
- If we know what the equation of $f(x)$ looks like (ex. $f(x) = x^3$), the we can often use calculus to obtain a maximum.
  - This becomes impractical quickly if the equation is nasty, or we are in a multivariate case.
- We need a way to find the maximum of the function without an equation of $f(x)$, and that only requires a small number of times we query $f(x)$. . .

Outline

What is a Gaussian Process (GP)?

- Main idea: Given a function $f(x)$, we create approximate functions that try and look like $f(x)$!
  - **Surrogate**: An "approximate" function.
  - We sample the value of the function at certain points, and use that to construct a surrogate model

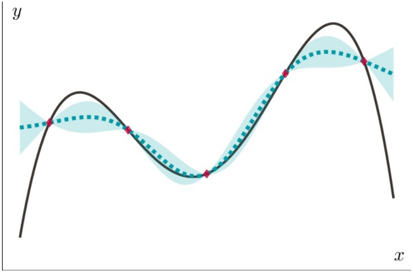


Figure: A surrogate (blue line) that approximates the actual function (black line) (image source)

What is a Gaussian Process (GP)?

- Main idea: Given a function $f(x)$, we create approximate functions that try and look like $f(x)$!
- The estimated function attempts to capture:
  - Frequency (how "wiggly" the function looks)
  - Amplitude (how big the "wiggles" are)

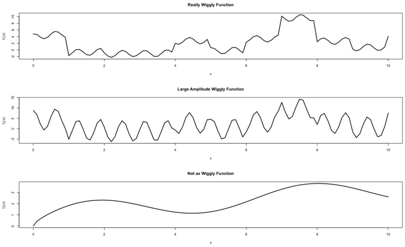


Figure: Several types of functions

What is a Gaussian Process (GP)?

- Main idea: Given a function $f(x)$, we create approximate functions that try and look like $f(x)$!
    - **Surrogate**: An "approximate" function.
    - We sample the value of the function at certain points, and use that to construct a surrogate model
- The estimated function attempts to capture:
    - Frequency (how "wiggly" the function looks)
    - Amplitude (how big the "wiggles" are)
- **Gaussian Process**: A collection of Random Variables (RVs) indexed by time or space such that every finite subset of these RVs follows a Normal Distribution.
    - We assume the equation of the function follows some sort of structure, with hyperparameters that the surrogate estimates.
- **Hyperparameters**:
    - τ: Controls scale (amplitude)
    - θ (Lengthscale): Controls the wiggleiness (frequency)

Advantages

- Speed
  - It can be much faster to find values of the estimated function, than the actual function.
- Flexibility
  - GPs can work quite well, even if the actual function is not friendly.
  - GPs extend quite naturally past univariate cases.

Outline

Optimization

Idea: What if we want to find the maximum of a function?

- GPs are faster, so it's easier to optimize over the surrogate model!

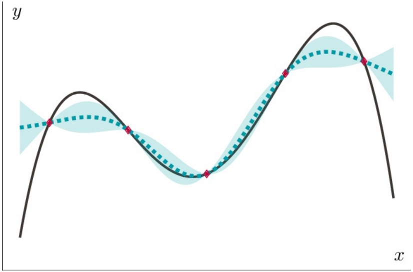


Figure: A surrogate (blue line) that approximates the actual function (black line). Also shown are uncertainty bands (blue region).

Optimization

Idea: What if we want to find the maximum of a function?

- GPs are faster, so it's easier to optimize over the surrogate model!
- Sequential methods add on samples to an existing sample set to find the maximum.
  - Where to sample next depends on the user.
    - One could sample at point with most uncertainty, or sample where surrogate's maximum is, etc.

Environmental Statistics

Note: *The author is not an expert on environmental statistics*

- Ex. Rainfall in US

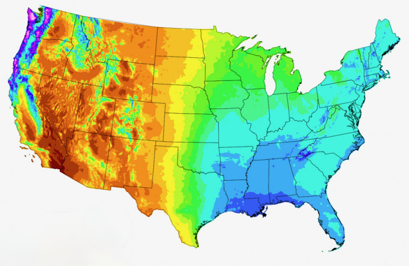


Figure: Precipitation map of the United states (image source).

Environmental Statistics

Note: *The author is not an expert on environmental statistics*

- Ex. Rainfall in US
- Ex. Pollution over time

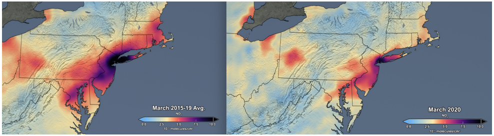


Figure: Pollution map of New England at two time points (image source).

Environmental Statistics

Note: *The author is not an expert on environmental statistics*

- Ex. Rainfall in US
- Ex. Pollution over time
- Ex. Whale populations around the world

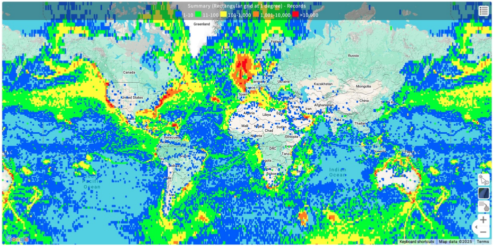


Figure: Ocean Biodiversity Information System (OBIS) data on whale sightings.