Machine Learning and Data Analytics
ME 5013- Fall 2019

The University of Texas at San Antonio™

<u>Lecture 07</u>

- Gaussian Process

Adel Alaeddini, PhD
Associate Professor of Mechanical Engineering
Advanced Data Engineering Lab
adel.alaeddini@utsa.edu

Why GPs?

- Here are some data points! What function did they come from?
  - I have *no idea*.
- Oh. Okay. Uh, you think this point is likely in the function too?
  - I have *no idea*.

Why GPs?

- You can’t get anywhere without making some assumptions
- GPs are a nice way of expressing this ‘prior on functions’ idea.
- Can do a bunch of cool stuff
  - Regression
  - Classification
  - Optimization

- Having some observed input-output pairs $(\mathbf{x}_{\mathrm{i}}, \mathrm{y}_i)$ where y_i might be corrupted by some noise ε_i

$$\mathrm{y}_i = f(\mathbf{x}_i) + \varepsilon_i \quad for \quad i = 1, \ldots\ldots, n$$

Scalar

Vector, i.e. $\mathbf{x}_i$ =$(x_i^{(1)}$,...., $x_i^{(p)})$

- ε_i is the additive independent identically distributed Gaussian noise with variance σ_n^2

- The underlying function f is not known (black-box function)

- The functional evaluation at the test (new) point $\mathbf{x} \subset X$ is denoted as f_* (or y_* or $f(\mathbf{x}_*)$)

Matrix, i.e. $X = (\mathbf{x}_1, \ldots, \mathbf{x}_n)$

Gaussian process is a collection of random variables, a finite number of which have a joint Gaussian distribution

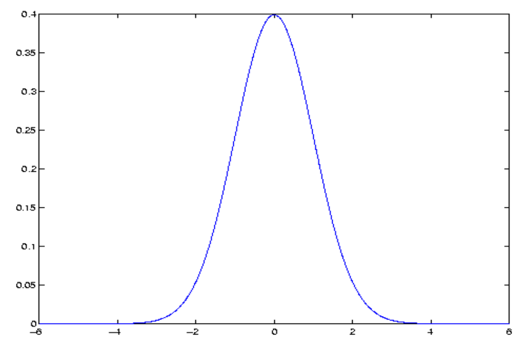
Input:

✓Training set : $\{(\mathbf{x}_i, y_i), i = 1,2,,..,n\}$ from $y_i = f(\mathbf{x}_i) + \varepsilon_i$

✓Test set X_* - Get predictions

Can be more than a single point $\mathbf{x}_{*1}, \ldots, \mathbf{x}_{*k}$

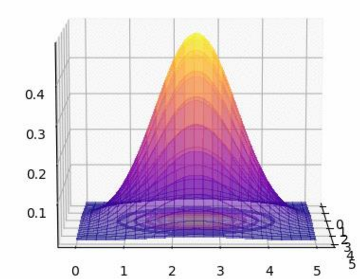
- According to the joint distribution of training outputs and test outputs we have,



$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(X,X) + \sigma_n^2 I & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right)$$

Kernel function between pair of points

- Squared Exponential Kernel is the most popular and highly used kernel



$$K(x,x') = \sigma_f^2 \exp(-\frac{1}{2l^2}(x - x')^2)$$

- By conditional distribution we get,

$$\bar{f}_* = \mathrm{E}(f_*|X, \mathbf{y}, X_*) = K(X, X_*)(K(X, X) + \sigma_n^2 I)^{-1}\mathbf{y}$$

$$\mathrm{cov}(\boldsymbol{f}_*) = [K(X_*, X_*) - K(X, X_*)\left[K(X, X) + \sigma_n^2 \boldsymbol{I}\right]^{-1} K(X, X_*)\big]$$

Procedure:

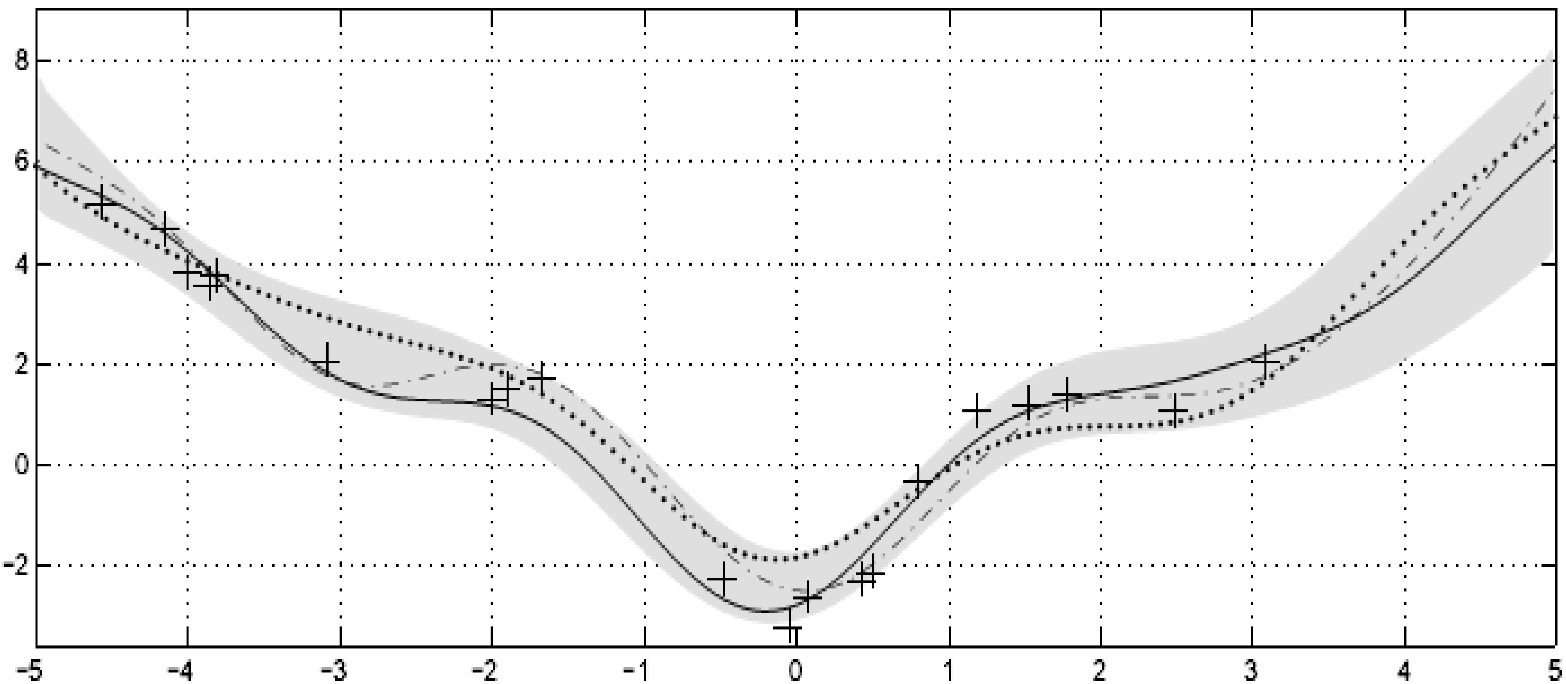
- The maximum likelihood estimate

$$J(a) = \frac{1}{2} a^T K a + \frac{1}{2\sigma_n^2}(y - Ka)^2$$

- First term is the ridge penalty term to prevent overfitting
- Second term is the standard loss function
- Minimizing above function, $\frac{\partial J}{\partial a} = 0$, gives $a = (K + \sigma_n^2 I)^{-1} y$

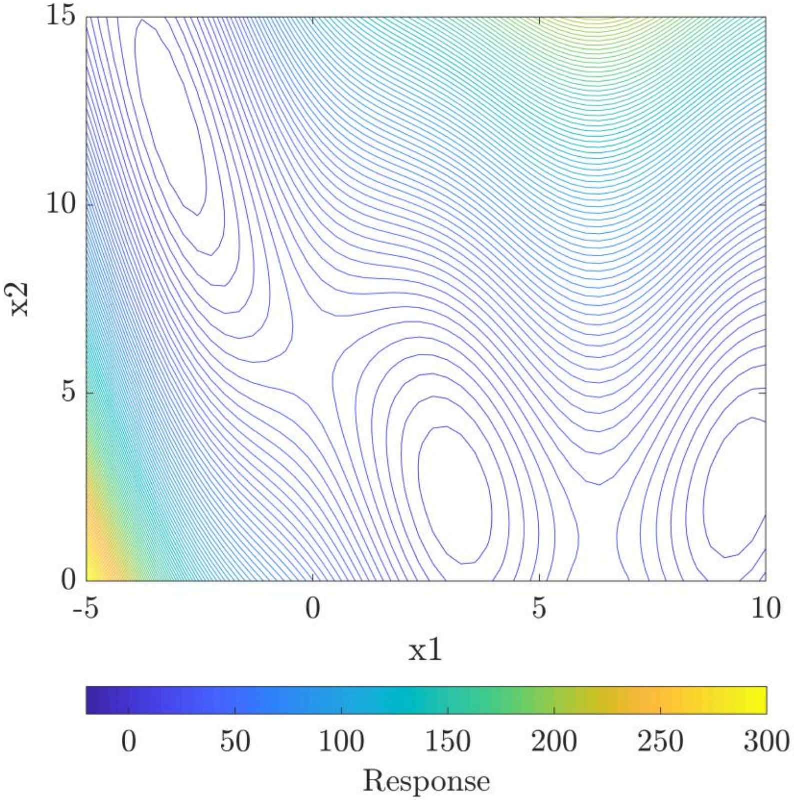
Example

- 20 Training data
- GP posterior
- noise level σ_n^2=0 and σ_n^2=7

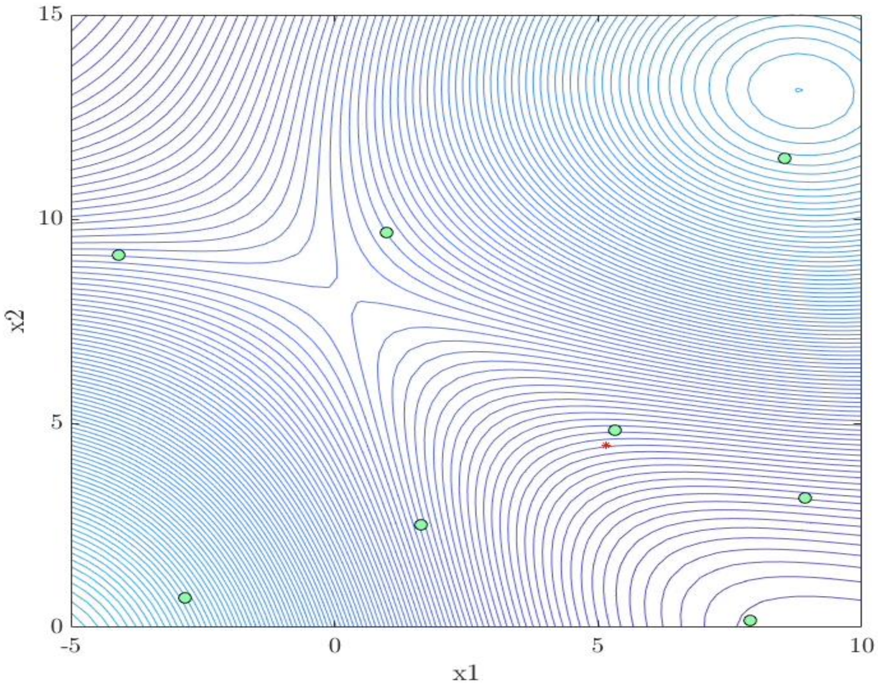


Example

True Contour



GP Estimated



Application: Image Reconstruction

Application: Image Reconstruction