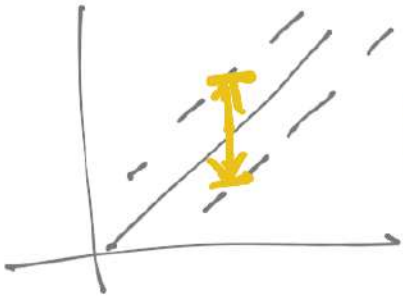
Advanced Machine Learning, 10 Sep 2019

Bayesian Optimization (GP Regression)

GP Regression

$$\begin{bmatrix} y & x \\ y_1 & x_1 \\ \vdots & \vdots \\ y_n & x_n \end{bmatrix}$$

$$y = \alpha + \beta x + \varepsilon$$



prediction within a confidence interval

Stochastic process

= Probability [German]

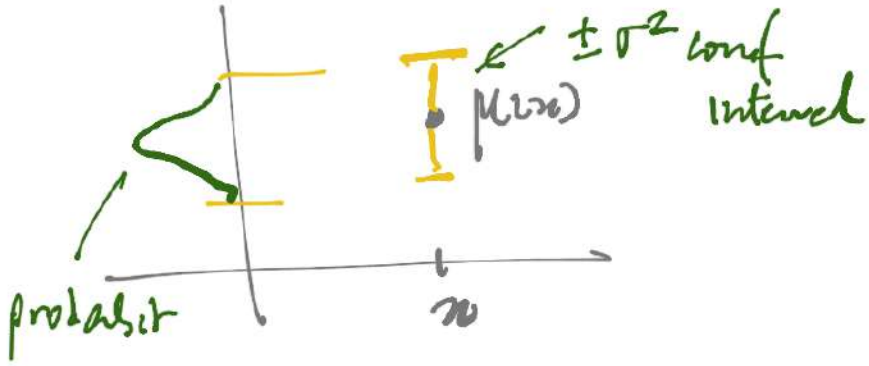
Why process

$$y = \mu(x) + \varepsilon \qquad \varepsilon \sim N(0, \sigma^2)$$

By additivity of $N(0, \sigma^2)$

$$y \sim N(\mu(x), \sigma^2)$$

x - index variable



$\mu(x)$ - mean is a function of x

σ^2 - constant

Can we make σ^2 a function of x

$$y \sim N(\mu(x), \Sigma(x))$$

Σ positive semi definite

Popular model

$$\Sigma = \sigma^2 \exp\{-\xi \|x - x'\|\}$$

One extra parameter ξ

Estimate using likelihood

Some exposures fix μ constant to model $\Sigma(x)$

Likelihood function

$$Y \sim N_n(\mu_{n\times n}, \Sigma_{n\times n})$$

$$\frac{1}{2|\Sigma|^{n/2}} \exp\left\{-\frac{1}{2}(y-\mu)^T \Sigma^{-1} (y-\mu)\right\}$$

parameters α, β

parameters σ, ρ inside Σ^{-1}

Each iteration of gradient descent requires inverting Σ

Bottleneck $\Rightarrow$ $n = 10$ or 20

Algorithm

m out of n bootstrap samples

Next time

Scales to $n = 10^6$