Lec 16

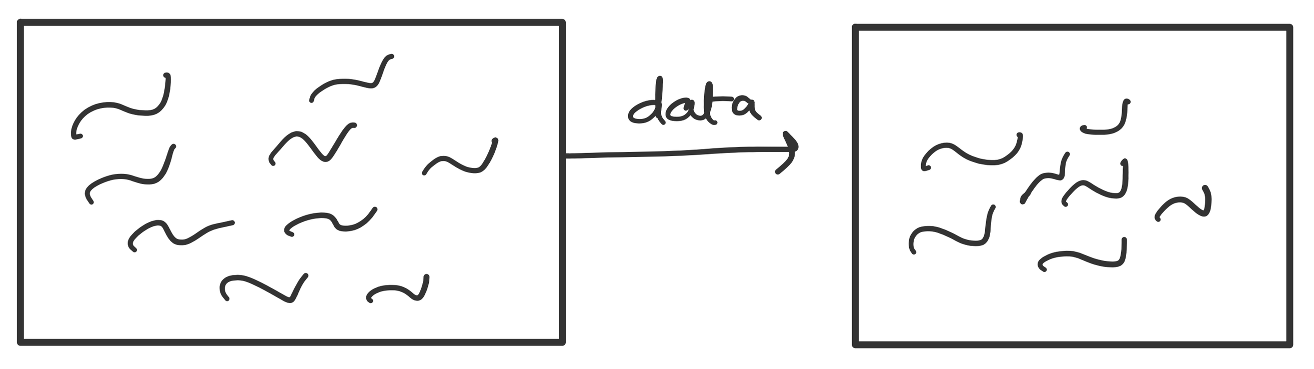
Wednesday, February 15, 2023 1:29 PM

Recap

- $y = w^T x + \varepsilon \xrightarrow{\text{Probabilistic}} P(y|x,w) \sim N(w^T x, \sigma^2 I)$

  $w \sim N(0,\Sigma_p)$, $\varepsilon \sim N(0,\sigma^2)$

- $y \sim GP(m(x), K(x,x'))$



$m(x)$, $K(x,x')$ $\longrightarrow$ indirectly specifying $\phi(x)$

$f(x) = \phi(x)^T w$

- prior knowledge, model selection, cross-validation

- $K(x,x') = \sigma_f^2 \exp\left(-\frac{1}{2\ell^2} d^2(x,x')\right)$ ← continuity

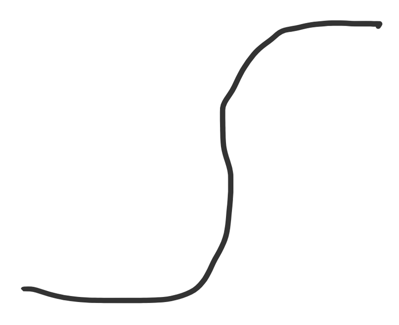
  $f(x), f(x') \leftarrow d(x,x')$

positive value $f(x) > 0$ $m(x) > 0$

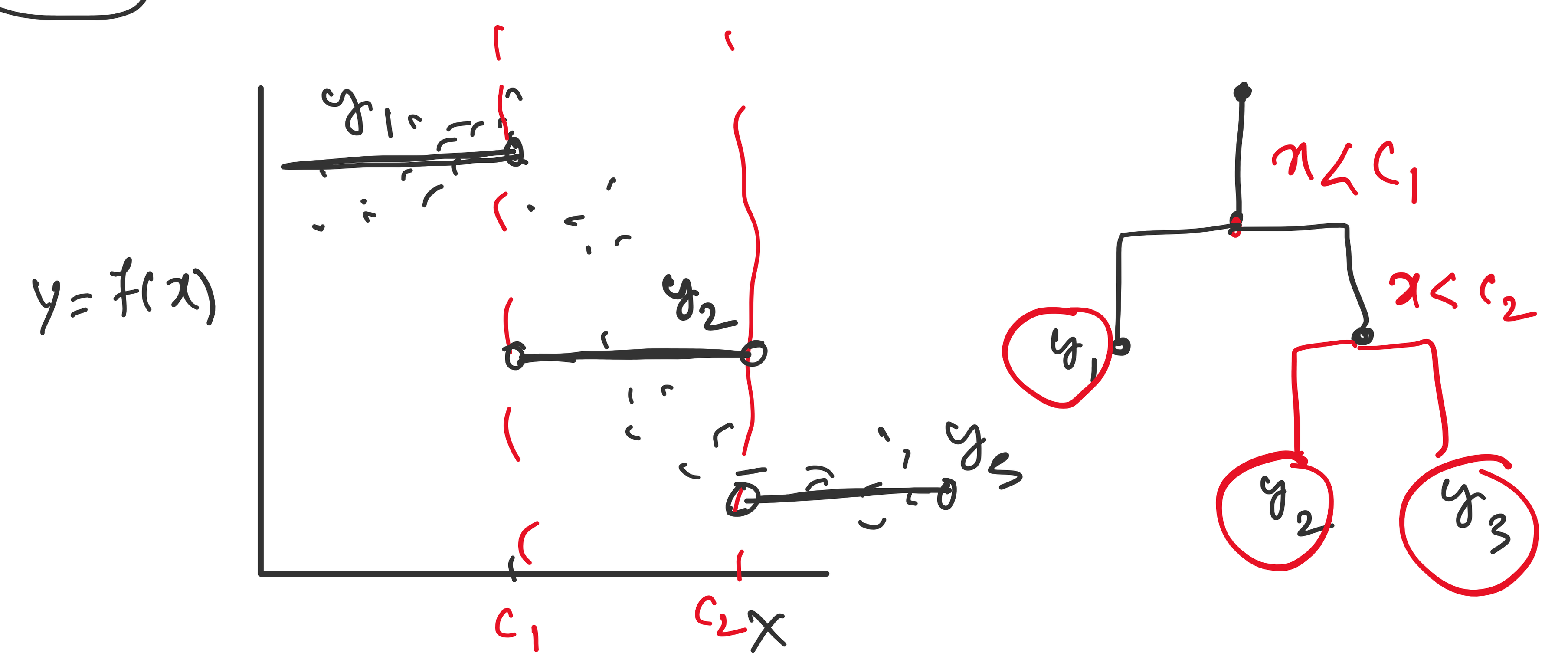
periodicity : cyclic voltammetry curves for n cycles.

$$K(x,x') = \exp\left(-\frac{2\sin^2(\pi d(x,x')/p)}{\ell^2}\right)$$

- you are inherently restricting your hypothesis class by pre-specifying kernels.

- 

Bayesian Additive Regression Trees



1 2 . . . K (# Tree functions)

→ 1 2 ⋮ B

$\rightarrow \bar{y} = \frac{1}{K}\sum f_i^b(x)$

$i \in (1,K)$

$b \in (1,B)$

variance : $Var(f_i^b(x))$

"ensemble" methods ← multiple functions to make a predictions

partial residual : $y_i - \underbrace{\sum_{k \neq k'} f_k^b(x_i)}$

change your k' tree at iteration b such that this is reduced.

Node prediction: $(y_1, y_2, y_3) \sim N(\mu, \sigma^2)$

a prior on μ ← split points you could use, what is your signal variance.

$\mu \sim N(0, \sigma_1^2)$

$\sigma \sim N(0, \sigma_2^2)$ (needs to be positive) log-normal Half-normal.