Bayesian Inference and Bandwidth

100 basis functions (RBFs), spaced between 0 and 1

Model 1: Bandwidth $h=1$  
Model 2: $h=0.01$

Data from model 1  
Data from model 2

$p(y|X, h=1) > p(y|X, h=0.01)$

$p(y|X, h=1)$  
$p(y|X, h=0.01)$

$\Rightarrow$ Marginal likelihood can be used for model choice

Can compare model on training data $\mathcal{D}$

Likelihood not suitable: $p(y|X, h=0.01, \hat{w})$ big for only 11 of 11 data sampled from model 1
Likelihood of all parameters:
\[ p(y \mid w, h, \ldots, X) \]

Marginal likelihood:
\[ \int p(y \mid w, h, \ldots, X) \cdot p(w \mid h, \ldots, X) \, dw \]
Bayesian Optimisation

System Performance

mean belief

x

2019 L13 (3)
Limitations of linear regression

Active selection

$\Rightarrow$ Make $x$ as big as possible with basis functions

- Certain that outputs will be 0 if we move away from basis functions
Gaussian processes
Very big Gaussian distribution
Gaussian process priors

\[ p(\tilde{f}) = \mathcal{N}(\tilde{f}; 0, \Sigma) \]

\[ \Sigma_{ij} = \text{cov}(\tilde{f}_i, \tilde{f}_j) \]

\[ = \mathbb{E}[\tilde{f}_i \tilde{f}_j] - \mathbb{E}[\tilde{f}_i] \mathbb{E}[\tilde{f}_j] \]

5-dim Gaussian
Covariance function / kernel function

Function prior \( f \sim \text{GP} \) with kernel \( k \):
For any subset of values \( \tilde{f} \):
\[
p(\tilde{f}) = \mathcal{N}(\tilde{f}; 0, K)
\]
where \( K_{ij} = k(x^{(i)}, x^{(j)}) \)

"kernel function" or "covariance function"

What functions can we use for \( k \)?

"Mercer kernels" / Positive definite kernels

\( \Rightarrow K \) positive semi-definite

Example:
\[
k(x^{(i)}, x^{(j)}) = \exp(-||x^{(i)} - x^{(j)}||^2)
\]

"Gaussian kernel" or "RBF kernel"
or "Squ. exponential kernel"

2019 L13⑦
Regression model with GP

Prior on functions $f \sim \text{GP}(k)$

Observation model:

$y_n \sim \mathcal{N}(\tilde{f}_n, \sigma_y^2)$

$\Rightarrow$ Likelihood $p(y_n | \tilde{f}) = \mathcal{N}(y_n; \tilde{f}_n, \sigma_y^2)$

for $n$ data points, product over $n = 1, \ldots, N$