Bayesian Inference & Bandwidth

Assume \( k = 100 \) RBF basis functions
Centers evenly spaced between 0 and 1
Samples from prior \( \mathbf{w} \sim N(0, \mathbf{I}_k) \)

---

**Typical observation**

Known noise level \( \sigma_y \)

\[
p(y \mid h=1) \gg p(y \mid h=0.01)
\]

---

**Marginal likelihood**

\[
p(\text{data} \mid \text{model}) = \int p(\text{data} \mid \mathbf{w}, \text{model}) \, p(\mathbf{w} \mid \text{model}) \, d\mathbf{w}
\]

Can compare models on training set.
Could also choose \( \sigma_y, \sigma_w \) in \( \mathbf{w} \sim N(0, \sigma_w^2 \mathbf{I}_k) \), ...
Likelihood of all your parameters

\[ p(y \mid w, h, \ldots, x) \]

Marginal likelihood:

\[ \int p(y \mid w, h, \ldots, x) \, p(w \mid h, \ldots, x) \, dw \]
Bayesian Optimization

How well system works
Limitations of linear regression

$$y = wx$$

Active selection

$$\Rightarrow$$ Make $x$ as big as you can.

$1, \phi_1, \phi_2, \phi_3 \rightarrow x$

Zero here
Gaussian Processes

Really big Gaussian distribution

\[ f = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_n \end{bmatrix} \]
Gaussian process prior

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

\[ \Sigma_{ij} = \text{cov}(f_i, f_j) \]

\[ = \mathbb{E}[f_i f_j] - \mathbb{E}[f_i] \mathbb{E}[f_j] \]

5-Dim Gaussian
Covariance / kernel function

Function prior $f \sim GP$ with kernel $k$

For any subset of values $\mathcal{F}$

$$p(\mathcal{F}) = N(\mathcal{F}; 0, K)$$

$$k_{ij} = k(x^{(i)}, x^{(j)})$$

kernel function
or covariance function

"Mercer kernels" / Positive definite functions

$\Rightarrow K$ semi-definite

Example

$$k(x^{(i)}, x^{(j)}) = \exp(-||x^{(i)} - x^{(j)}||^2)$$

(an RBF)
Regression model

Prior on functions $f \sim GP(k)$

Observation model:

$$y_n \sim N(f_n, \sigma^2_y)$$

$\Rightarrow$ Likelihood $p(y_n \mid f) = N(y_n; f_n, \sigma^2_y)$

Product over $n=1 \ldots N$

Posterior and Predictions

$$p(f \ast 1 y, X)$$

$$p(y \ast 1 y, X)$$