Machine Learning 1
Lecture 5.1 - Supervised Learning
Bayesian Linear Regression - The Equivalent Kernel

Erik Bekkers

(Bishop 3.3.3)
Equivalent Kernel Formulation

- Predictive distribution
  
  \[
p(t' | x', X, t, \alpha, \beta) = \int p(t' | x', w, \beta) p(w | X, t, \alpha, \beta) \, dw
  \]
  
  \[
  = \mathcal{N}(t' | m_N^T \phi(x'), \sigma_N^2(x'))
  \]

  \[
m_N = \beta S_N \Phi^T t \\
  \sigma_N^2(x') = \frac{1}{\beta} + \phi(x')^T S_N \phi(x')
  \]

- Predictive mean:

  \[
y(x', m_N) = \phi(x')^T m_N
  \]

  \[
  = \beta \phi(x')^T S_N \Phi^T t = \beta \phi(x')^T S_N \sum_{n=1}^{N} \phi(x_n) b_n
  \]

  \[
  = \sum_{n=1}^{N} \beta \phi(x')^T S_N \phi(x_n) b_n = \sum_{n=1}^{N} k(x', x_n) b_n
  \]

- Equivalent kernel

  \[
k(x', x) = \beta \phi(x')^T S_N \phi(x_n)
  \]
Equivalent kernel for Gaussian Basis Functions

Figure: Equivalent kernel $k(x', x)$ (Bishop 3.10)

- Localized kernel
  \[ k(x', x) = \beta \phi(x')^T S_N \phi(x) \]
- Predictive mean
  \[ y(x', m_N) = \sum_{n=1}^{N} k(x', x_n) t_n \]
- Training points $x_n$ close to $x'$ contribute more!
- Covariance of between predictions:
  \[ \text{cov}[t_1, t_2 | x_1, x_2] = \text{cov}_w [y(x_1, w), y(x_2, w)] \]
  \[ = \text{cov}_w [\phi(x_1)^T w, w^T \phi(x_2)] = \mathbb{E}_w [\phi(x_1)^T w w^T \phi(x_2)] - \mathbb{E}_w [\phi(x_1)^T w] \mathbb{E}_w [w^T \phi(x_2)] \]
  \[ = \phi(x_1)^T (\mathbb{E}_w [w w^T] - \mathbb{E}_w [w] \mathbb{E}_w [w^T]) \phi(x_2) = \phi(x_1)^T \text{cov}_w (\phi(x_2)) + \phi(x_1)^T \text{cov}_w (\phi(x_2)) \]
  \[ p(w | X, t, \alpha, \beta) = \mathcal{N}(w | m_N, S_N) \]
  \[ = \frac{1}{\sqrt{(2\pi)^d |S_N|}} \exp \left( -\frac{1}{2} (w - m_N)^T S_N (w - m_N) \right) \]
  \[ \mathbb{E}[t' | x', w] = y(x', w) \]