CS/CNS/EE 253 – Advanced Topics in Machine Learning
Problem Set 3

Handed out:	24 Feb 2010
Due:	8 Mar 2010

---

1 Reproducing Kernel Hilbert Spaces

Let $\mathcal{X}$ be an index set, and k, k' two positive definite kernel functions defined on $\mathcal{X} \times \mathcal{X}$. Let $\mathcal{H}_k$ and $\mathcal{H}_{k'}$ be the Reproducing Kernel Hilbert Spaces associated with k and k' respectively. Show that if $\mathcal{H}_k = \mathcal{H}_{k'}$ then $k = k'$, i.e., for all $x, x' \in \mathcal{X}$, $k(x, x') = k'(x, x')$.

2 Gaussian Processes and Bayesian Linear Regression

The regression problem involves estimating the functional dependence between an input variable $\mathbf{x}$ in $\mathbb{R}^d$ and an output variable y in $\mathbb{R}$. We assume the relationship

$$y(\mathbf{x}) = \sum_{j=1}^{M} w_j \phi_j(\mathbf{x}) + \epsilon = \mathbf{w}^T \mathbf{\Phi}(\mathbf{x}) + \epsilon, \tag{1}$$

where $\mathbf{w}^T \mathbf{\Phi}(\mathbf{x})$ is a linear combination of M predefined nonlinear basis functions $\phi_j(\mathbf{x})$ with input in $\mathbb{R}^d$ and output in $\mathbb{R}$. The observations are additively corrupted by i.i.d. noise with normal distribution

$$\epsilon \sim N(0, \sigma_n^2) \tag{2}$$

which has zero mean and variance σ_n^2.

Our goal is to estimate the weights w_j given a training set consisting of pairs $(\mathbf{x_1}, y_1), \ldots, (\mathbf{x_n}, y_n)$. We use a multivariate normal distribution as prior on the weights

$$\mathbf{w} \sim N(0, \mathbf{\Sigma}_w), \tag{3}$$

with zero mean and M-by-M sized covariance matrix $\mathbf{\Sigma}_w$.

The goal of this problem is to show that the Bayesian linear regression defined above is an example of a Gaussian process. Recall that a Gaussian Process is a probability measure over $y(\mathbf{x})$ defined by a mean function $\mu(\mathbf{x}) = E[y(\mathbf{x})]$ and a covariance kernel $k(\mathbf{x}, \mathbf{x}') = E[(y(\mathbf{x}) - \mu(\mathbf{x}))(y(\mathbf{x}') - \mu(\mathbf{x}'))]$. We can write this as

$$y(\mathbf{x}) \sim \mathrm{GP}(\mu(\mathbf{x}), k(\mathbf{x}, \mathbf{x}')). \tag{4}$$

1. Show that the Bayesian linear regression functions defined above have mean function

$$\mu(\mathbf{x}) = 0 \tag{5}$$

and covariance function

$$k(\mathbf{x}, \mathbf{x}') = \mathbf{\Phi}(\mathbf{x})^T \mathbf{\Sigma}_w \mathbf{\Phi}(\mathbf{x}') + \sigma_n^2 \delta(\mathbf{x}, \mathbf{x}'), \tag{6}$$

where $\delta(\mathbf{x}, \mathbf{x}') = 1$ if $\mathbf{x} = \mathbf{x}'$ and zero otherwise.

2. Prove that the covariance function (eq. 6) is a valid kernel function. That is, prove that it is symmetric and positive semi-definite.

3. Derive an expression for

$$P(y'|\mathbf{x}', (\mathbf{x_1}, y_1), \ldots, (\mathbf{x_n}, y_n)) \tag{7}$$

which is the predictive distribution of the output variable y' associated with test point $\mathbf{x}'$ given that we have observed a training data set $(\mathbf{x_1}, y_1), \ldots, (\mathbf{x_n}, y_n)$.

4. Derive an expression for the 95-th percentile of the predictive distribution of y'.

5. Use $[1\ x\ x^2\ x^3]^T$ as your basis functions, $\mathbf{\Sigma}_w$ equal to the identity matrix, and $\sigma_n = 0.1$. Implement Bayesian linear regression and run it on the following data set: $(x_1 = 1, y_1 = 0.5), (x_2 = 2, y_2 = 1.6), (x_3 = 3, y_3 = 1.1), (x_4 = 4, y_4 = 3), (x_5 = 5, y_5 = 4.2)$. What are the predictive distributions associated with the following test points: $x' = \{1.5, 2.5, 3.5, 4.5, 5.5\}$? What is the value associated with the 95-th percentile for each test point? Please submit your code.