Assignment 4: Kernels and Neural Networks

CS489/698 – Winter 2010

Out: March 18, 2010
Due: April 1, 2010

Be sure to include your name and student number with your assignment.

1. **[15 pts]** Show that the Gaussian kernel $k(x, x') = exp(-||x - x'||^2/2\sigma^2)$ can be expressed as the inner product of an infinite-dimensional feature space. Hint: use the following expansion and show that the middle factor further expands as a power series:

$$k(x, x') = e^{-x^T x/2\sigma^2} e^{x^T x'/\sigma^2} e^{-(x')^T x'/2\sigma^2}$$

2. **[15 pts]** Consider a two-layer neural network of the form

$$y_k(x, w) = \sigma(\sum_j w_{jk}^{(2)} h(\sum_i w_{ij}^{(1)} x_i + w_{0j}^{(1)}) + w_{0k}^{(2)})$$

   in which the hidden unit nonlinear activation functions $h(\cdot)$ are given by logistic sigmoid functions of the form

$$\sigma(a) = \frac{1}{1 + e^{-a}}.$$

   Show that there exists an equivalent network, which computes exactly the same function, but with hidden unit activation functions given by

$$tanh(a) = \frac{e^a - e^{-a}}{e^a + e^{-a}}.$$

   Hint: first find the relation between $\sigma(a)$ and $tanh(a)$ and then show that the paramneters of the two neural networks differ by linear transformations.

3. **[20 pts]** For this question, you will develop a dual formulation of the perceptron learning algorithm. Using the perceptron learning rule

$$w^{t+1} = \begin{cases} w^t + y_n \phi(x_n) & \text{if } y_n w^T \phi(x_n) \leq 0 \\ w^t & \text{otherwise} \end{cases}$$

   show that the learned weight vector w can be written as a linear combination of the vectors $y_n \phi(x_n)$ where $y_n \in \{-1, +1\}$. Denote the coefficients of this linear combination by α_n.

   (a) **[10 pts]** Derive a formulation of the perceptron learning rule in terms of α_n. Show that the feature vector $\phi(x)$ enters only in the form of the kernel function $k(x, x') = \phi(x)^T \phi(x')$.

   (b) **[10 pts]** Derive a formulation of the predictive learning rule

$$y = \begin{cases} 1 & \text{if } w^T \phi(x) > 0 \\ -1 & \text{otherwise} \end{cases}$$

4. **[50 pts]** Non-linear models for classication.

   Implement the following two classification algorithms. A dataset will be posted on the course web page. The input space is continuous (i.e., $X = \Re^d$), while the output space is categorical (i.e., $Y = \{C_1, C_2\}$).

   (a) **[25 pts]** Kernel perceptron: using the learning rule that you derived in Question 3, learn the coefficients a_n for the following kernels:
   - Identity: $k(x, x') = x^T x'$
   - Gaussian: $k(x, x') = e^{-||x-x'||^2/2\sigma^2}$
   - Polynomial: $k(x, x') = (x^T x' + 1)^d$ where d is the degree of the polynomial

   (b) **[25 pts]** Neural network: learn the weights of a two-layer neural network with a sigmoid activation function for the hidden and output nodes.