Assignment 2

Due: Monday June 15, 2015

Instructions: Please complete the problems independently and send the softcopy of your answers to `changhong@ict.ac.cn`.

1. (20 points) Show that, irrespective of the dimensionality of the data space, a data set consisting of just two data points (call them $\mathbf{x}^{(1)}$ and $x^{(2)}$, one from each class, is sufficient to determine the maximum-margin hyperplane. Fully explain your answer, including giving an explicit formula for the solution to the hard margin SVM (i.e., $\mathbf{w}$) as a function of $\mathbf{x}^{(1)}$ and $x^{(2)}$.

2. (20 points) Problem 3.11 in Bishop's book

   We have seen that, as the size of a data set increases, the uncertainty associated with the posterior distribution over model parameters decreases. Make use of the matrix identity

   $$(\mathbf{M} + \mathbf{v}\mathbf{v}^T)^{-1} = \mathbf{M}^{-1} - \frac{(\mathbf{M}^{-1}\mathbf{v})(\mathbf{v}^T\mathbf{M}^{-1})}{1 + \mathbf{v}^T\mathbf{M}^{-1}\mathbf{v}}$$

   to show that the uncertainty $\sigma_N^2(\mathbf{x})$ associated with the linear regression function given by (3.59) satisfies

   $$\sigma_{N+1}^2(\mathbf{x}) \leq \sigma_N^2(\mathbf{x}).$$

3. (20 points) Problem 5.28 in Bishop's book

   Consider a neural network, such as the convolutional network discussed in Section 5.5.6, in which multiple weights are constrained to have the same value. Discuss how the standard backpropagation algorithm must be modified in order to ensure that such constraints are satisfied when evaluating the derivatives of an error function with respect to the adjustable parameters in the network.

4. (20 points) Problem 6.11 in Bishop's book

   By making use of the expansion (6.25), and then expanding the middle factor as a power series, show that the Gaussian kernel (6.23) can be expressed as the inner product of an infinite-dimensional feature vector.

5. (20 points) Hidden Markov Models

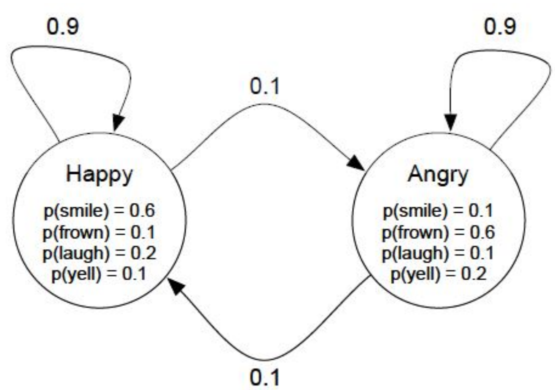

Figure 1: A Hidden Markov model

Andy lives a simple life. Some days he is Angry and some days he is Happy. But he hides his emotional state, and so all we can observe is whether he smiles, frowns, laughs, or yells. Andy's best friend is utterly confused about whether Andy is actually happy or angry and decides to model his emotional state using a hidden Markov model.

Let $X_d \in \{Happy, Angry\}$ denote Andy's emotional state on day d, and $\mathbf{Y}_d \in \{smile, frown, lauth, yell\}$ denote the observation made about Andy on day d. Assume that on day 1 Andy is in the Happy state, i.e. $X_1 = Happy$. Furthermore, assume that Andy transitions between states exactly once per day according to the distribution labeled in Fig. 1. The observation distributions for Andy's Happy and Angry state are given in Fig. 1 as well.

(1) $P(X_2 = Happy)$?

(2) $P(Y_2 = frown)$?

(3) $P(X_2 = Happy|Y_2 = frown)$?

(4) $P(Y_{80} = yell)$?

(5) Assume that $Y_1 = Y_2 = Y_3 = Y_4 = Y_5 = frown$. What is the most likely sequence of the states? That is, compute the MAP assignment $\arg\max_{x_1,\ldots,x_5} P(X_1 = x_1, \ldots, X_5 = x_5 | Y_1 = Y_2 = Y_3 = Y_4 = Y_5 = frown)$.