CS60050: Machine Learning

Endsemester Exam Spring 2012

Time =3 hrs, Total Marks = 100, Answer all questions. In each question show all the intermediate steps.

---

1. Suppose we have four training examples in two dimensions. Positive examples at $X_1 = [\,0,\ 0\,]$, $X_2 = [\,2,\ 2\,]$, and negative examples at $X_3 = [\,h,\ 1\,]$, $X_4 = [\,0,\ 3\,]$, where $h >= 0$ is a parameter. [5 X 3]

(a) How large can h be so that the training points are still linearly separable?

(b) Does the orientation of the decision boundary change as a function of h when the points are still linearly separable? Explain.

(c) What is the margin achieved by the maximum margin hyperplane as a function of h? Explain.

2. A set of n points (each point denoted by X) are partitioned into c disjoint subsets, $D_1, D_2, \ldots, D_c$. Let m_i be the mean of the points in subset D_i. The sum of squared error of the partition may be defined as

$$J_e = \sum_{i=1}^{c} \sum_{X \in D_i} \| X - m_i \|^2$$

(a) Consider a set of $n = 2k + 1$ points, k of which coincide at $X = -2$, k at $X = 0$, and one at $X = a > 0$.

(i) Show that the partitioning that minimizes J_e groups the points at $X = 0$, with the one at $X = a$, if $a^2 < 2(k+1)$.

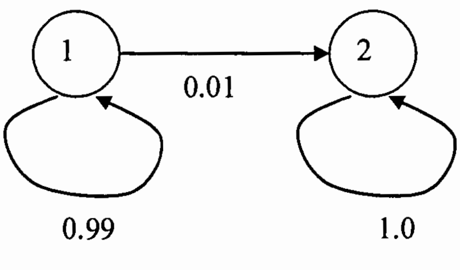
(ii) What is the optimal grouping for $a^2 > 2(k+1)$? [5 + 5]

(b) Consider a hierarchical clustering procedure in which clusters are merged so as to produce the smallest increase in sum squared error J_e at each step. If the ith cluster contains n_i points with mean m_i, show that the smallest increase results from merging the pair of clusters for which [10]

$$\frac{n_i n_j}{n_i + n_j} \| m_i - m_j \|^2 \text{ is minimum.}$$

3. Consider the two state HMM shown below. The transition probabilities are shown on the arcs. The output distribution for each set is defined over $\{1, 2, 3, 4\}$ and is given in the table below. The HMM is equally likely to start from either of the two states. [5 X 4]

State	*1*	*2*
P(X=1)	0.0	0.1
P(X=2)	0.199	0.0
P(X=3)	0.8	0.7
P(X=4)	0.001	0.2



(a) Give an example of an output sequence of length two which cannot be generated by the HMM.

(b) We generated a sequence of 6867^{2012} observations from the HMM and found that the last observation was 3. What is the most likely hidden state corresponding to the last observation?

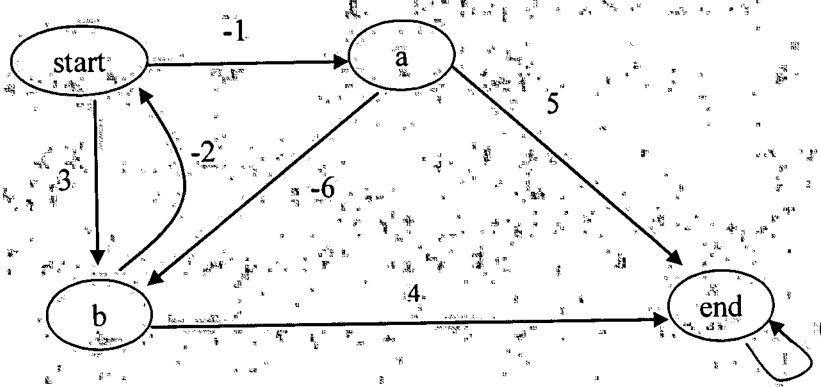
(c) Consider an output sequence 3, 3. What is the most likely sequence of hidden states corresponding to this observation?

(d) Now consider an output sequence 3, 3, 4. What are the first two states of the most likely hidden state sequence?

4. Write short notes on the following: [5 X 5]

(a) Parzen window estimate, (b) Kullback-Liebler divergence, (c) Mahalanobis distance,
(d) Karhunen-Loeve transform, (e) Vapnik-Chervonenkis dimension.

5. Consider the reinforcement environment drawn below (let $\gamma = 0.9$). The numbers on the arcs indicate the immediate rewards. [5 X 4]



(a) Show the action at each state representing the optimal policy.
(b) Compute the V^* values at every state.
(c) Assume we use a *Q table* for this task and initialize all of its entries to *2*. A learner then follows the path *start* $\rightarrow$ *b* $\rightarrow$ *end*. Show the updated Q values on the graph above.
(d) Starting with the Q values updated above, again follow the path *start* $\rightarrow$ *b* $\rightarrow$ *end*, and show, on the graph above, the Q values that have changed.

------------------- **BEST WISHES** -------------------