CS534 — Written Assignment 3 — Due May 16th

Written assignment
Due - May 16th In class

- MAP estimation. Consider the problem of linear regression. We are given a set of observed data points $(X_i, t_i) : i = 1, \cdots, N$, where X is the input vector, and t is the target output. The goal is to estimate a set of linear coefficients W such that t can be predicted by $W^T X$. In particular, we assume that $t|X \sim N(W^T X, \sigma^2)$. Now we further assume that each coefficient w_i has a prior distribution $N(0; \alpha^{-1})$. Please write down the posterior function of W, and show that maximizing this posterior is equivalent to minimizing the least square objective with a L_2 regularization term.

- Boosting. Please show that in each iteration of Adaboost, the weighted error of h_i on the updated weights D_{i+1} is exactly 50%. In other words, $\sum_{j=1}^{N} D_{i+1}(j) I(h_i(X_j) \neq y_j) = 50\%$.

- PAC learnability. Consider the concept class C of all conjunctions (allowing negations) over n boolean features. Prove that this concept class is PAC learnable.

- VC dimension. Consider the hypothesis space H_r = the set of all rectangles in the 2-d (x, y) plane. That is, $H = \{((a < x < b) \wedge (c < y < d)) \mid a, b, c, d \in \Re\}$. What is the VC dimension of H_r. Provide a proof to your claim.

- Consider the class C of concepts of the form $(a \leq x \leq b) \wedge (c \leq y \leq d)$, where a, b, c, and d are integers in the interval $[0, 99]$. Note that each concept in this class corresponds to a rectangle with integer-valued boundaries on a portion of the (x, y) plane. Hint: Given a region in the plane bounded by the points $(0, 0)$ and $(n-1, n-1)$, the number of distinct rectangles with integer-valued boundaries within this region is $\left(\frac{n(n-1)}{2}\right)^2$.

  (a) Give an upper bound on the number of randomly drawn training examples sufficient to assure that for any target concept c in C, any consistent learner using $H = C$ will, with probability 95%, output a hypothesis with error at most 0.15.

  (b) Now suppose the rectangle boundaries a, b, c, and d take on *real* values instead of integer values. Update your answer to the first part of this question.