**Homework #3**
TA in charge: Han-Hsing Tu

RELEASE DATE: 10/23/2008

DUE DATE: 10/30/2008, 4:00 pm IN CLASS

TA SESSION: 10/29/2008, noon to 2:00 pm IN R106

*Unless granted by the instructor in advance, you must turn in a hard copy of your solutions (without the source code) for all problems. For problems marked with (*), please follow the guidelines on the course website and upload your source code to designated places.*

*Discussions on course materials and homework solutions are encouraged. But you should write the final solutions alone and understand them fully. Books, notes, and Internet resources can be consulted, but not copied from.*

*Since everyone needs to write the final solutions alone, there is absolutely no need to lend your homework solutions and/or source codes to your classmates at any time. In order to maximize the level of fairness in this class, lending and borrowing homework solutions are both regarded as dishonest behaviors and will be punished according to the honesty policy.*

*You should write your solutions in English with the common math notations introduced in class or in the problems. We do not accept solutions written in any other languages.*

## 3.1 Growth Function and VC Dimension

(1) (20%) Find the growth function $L(N)$ and the VC dimension $D$ for the "interval" learning model $\{g_{\alpha,\beta}\}$, where $g_{\alpha,\beta}\colon \mathbb{R} \to \{-1,+1\}$. For any input $x \in \mathbb{R}$, $g_{\alpha,\beta}(x) = +1$ if $x$ lies within $[\alpha,\beta]$, and $-1$ otherwise. Mathematically verify that $L(N) \le N^D + 1$.

(2) (15%) Let $D$ be the VC dimension of the "triangle" learning model $\{g_T\}$, where $g_T\colon \mathbb{R}^2 \to \{-1,+1\}$. For any input $x \in \mathbb{R}^2$, $g_T(x) = +1$ if $x$ lies within a triangle $T$, and $-1$ otherwise. Prove that $D \ge 7$. (*Hint: put 7 points on a circle.*)

(3) (15%) Let $D$ be the VC dimension of the 'convex polygon' learning model $\{g_C\}$, where $g_C\colon \mathbb{R}^2 \to \{-1,+1\}$. For any input $x \in \mathbb{R}^2$, $g_C(x) = +1$ if $x$ lies within a convex polygon $C$, and $-1$ otherwise. Prove that $D$ is infinite. (*Hint: put $N$ points on a circle.*)

## 3.2 VC Dimension of Perceptron

In this problem, we consider the set $G_P$ of all possible perceptrons $g_{w,\theta}\colon \mathbb{R}^d \to \{-1,+1\}$. Each perceptron is of the form $g_{w,\theta}(x) = \operatorname{sign}\Big(\langle w, x\rangle - \theta\Big)$. We will prove that the VC dimension of $G_P$ is exactly $(d+1)$ by showing that

(1) (25%) There are $(d+1)$ points on which all $2^{d+1}$ label patterns can be produced from $G_P$. *(Hint: construct an invertible matrix using $(d+1)$ points.)*

(2) (25%) There are no $(d+2)$ points on which all $2^{d+2}$ label patterns can be produced from $G_P$. *(Hint: use the fact that any $(d+2)$ vectors of dimension $(d+1)$ have to be linearly dependent.)*

## 3.3 Experiment with 1-Nearest Neighbor (*)

(1) (20%) Implement the 1-Nearest algorithm $A$: For a given training set $Z = \Big\{(x_n, y_n)\Big\}_{n=1}^{N}$, the algorithm "returns" a decision function $g = A(Z)$, where $g$ classifies any given input $x$ by

$$g(x) = y_n, \text{ where } x_n \text{ is the nearest neighbor to } x \text{ (ties arbitrarily broken)}.$$

Run the algorithm on the following set for training:

`http://www.csie.ntu.edu.tw/~htlin/course/ml08fall/data/hw3_train.dat`

and the following set for testing:

`http://www.csie.ntu.edu.tw/~htlin/course/ml08fall/data/hw3_test.dat`

Report $\nu\Big(A(Z)\Big)$ and $\hat{\pi}\Big(A(Z)\Big)$.

(2) (10%) Randomly separate $Z$ to a base set $Z_b$ of size 80 and a validation set $Z_v$ of size 20. Compute the validation error $\nu_v\Big(A(Z_b)\Big)$. *(Note: see homework 2 for the definition of $\nu_v$)*

(3) (20%) Let $\nu_c\Big(A, \kappa\Big)$ denote the $\kappa$-fold cross-validation error of $A$. That is, $\nu_c\Big(A, \kappa\Big)$ is computed from the following procedure:

(a) Randomly separate $Z$ to disjoint sets $Z_1, Z_2, \cdots, Z_\kappa$, where each set is of (roughly) equal size.

(b) For $i = 1, 2, \cdots, \kappa$, run $A$ using $Z_b = Z - Z_i$ and $Z_v = Z_i$, and compute $\nu_i = \nu_v\Big(A(Z_b)\Big)$.

(c) Obtain the average error ($\kappa$-fold cross-validation error):

$$\nu_c\Big(A, \kappa\Big) = \frac{1}{|Z|}\sum_{i=1}^{\kappa} |Z_i| \cdot \nu_i.$$

Plot $\nu_c\Big(A, \kappa\Big)$ as a function of $\kappa$ for $\kappa = 2, 4, 5, 10, 20, 25, 50, 100$. Compare $\nu_c\Big(A, \kappa\Big)$ with the values of $\nu_v\Big(A(Z_b)\Big)$ and $\hat{\pi}\Big(A(Z)\Big)$.

## 3.4 Experiment with $K$-Nearest Neighbor (*)

(1) (25%) Implement the $K$-Nearest algorithm $A_K$: For a given training set $Z = \Big\{(x_n, y_n)\Big\}_{n=1}^{N}$, the algorithm "returns" a decision function $g = A_K(Z)$ that classifies any given input $x$ by

$$g(x) = \text{sign}\left(\sum_{k=1}^{K} y_{n_k}\right), \text{ where } x_{n_k} \text{ is the } k\text{-th nearest neighbor to } x \text{ (ties arbitrarily broken).}$$

Run the algorithm on the following set for training:

`http://www.csie.ntu.edu.tw/~htlin/course/ml08fall/data/hw3_train.dat`

and the following set for testing:

`http://www.csie.ntu.edu.tw/~htlin/course/ml08fall/data/hw3_test.dat`

Plot $\nu\Big(A_K(Z)\Big)$ and $\hat{\pi}\Big(A_K(Z)\Big)$ as a function of $K$ on the same figure for $K = 1, 3, 5, 7, 9, 11, 13, 15$. *(Note: $A_1$ is simply $A$ in Problem 2.3)*

(2) (25%) Plot $\nu_c\Big(A_K, 10\Big)$ as a function of $K$ for $K = 1, 3, 5, 7, 9, 11, 13, 15$. Compare $\nu_c\Big(A_K, 10\Big)$ with the values of $\hat{\pi}\Big(A_K(Z)\Big)$.

## 3.5 Mysterious $B$-function Leads to Bonus

Recall that we proved $B(N, M) \le \sum_{m=0}^{M-1} C(N, m)$ in class.

(1) (Bonus 10%) Prove the following inequality: $B(N, M) \ge \sum_{m=0}^{M-1} C(N, m)$.

Thus, $B(N, M) = \sum_{m=0}^{M-1} C(N, m)$.