B565 HW5

1. Create a collection of 50 two-dimensional labeled data points $\{(x_1, y_1)\ldots,(x_n, y_n)\}$ with $x_i \in [-\pi, +\pi] \times [-\pi, +\pi]$ and $y_i \in \{+1, -1\}$ such that

$$y_i(w_0 + w_1 x_{i1} + w_2 x_{i2} + w_3 \cos(x_{i1}) + w_4 \sin(x_{i1})) > 0$$

for $i = 1, \ldots, n$ for randomly chosen parameters (independent $N(0, 1)$) $w_0, \ldots, w_4$.

   (a) Plot the data using different symbols for the two classes.

   (b) Find the optimal parameters $\hat{w}_0, \ldots, \hat{w}_4$ that maximize the margin generated by the level set

$$\{(x_1, x_2) : \hat{w}_0 + \hat{w}_1 x_{i1} + \hat{w}_2 x_{i2} + \hat{w}_3 \cos(x_{i1}) + \hat{w}_4 \sin(x_{i1}) = 0\}$$

   (c) Add the decision boundary to your points by drawing the resulting level set.

2. Suppose $\hat{y}_1, \ldots, \hat{y}_M$ are classifiers that estimate a binary class label y. Suppose that the classifiers give the correct response with probability $p = .55$, and that the classifiers are *independent*.

   (a) Construct a sample of $n = 1000$ binary variables (0-1) y_i, $i = 1, \ldots, n$ representing both classes approximately equally (you should choose these randomly). For each y_i generate $M = 100$ independent binary variables, $\hat{y}_{i1}, \ldots, \hat{y}_{iM}$, (the classifiers) that correctly identify the class, y_i, with probability p,

   (b) For each example y_i construct the voting estimate $\hat{y}_i$ by voting among the $\hat{y}_{i1}, \ldots, \hat{y}_{iM}$.

   (c) Estimate the probability of correct classification of your "voting" classifier by counting the number of correctly classified examples among the $i = 1, \ldots, n$. The result should show significant improvement over the baseline error rate of $1 - p$.

   (d) Perform an analogous experiment with 3 different classes where your classifiers show a slight preference for the true class ($p = 1/3 + .01$) over the wrong classes ($p = 1/3 - .005$). Determine your final classifier by voting among the n independent weak classifiers. Your voting classifier should demonstrate a substantial improvement over the original classifiers.

3. Suppose we have a binary classification problem where the data $(x_1, y_1)\ldots,(x_n, y_n)$ have $x_i \in \Re$, $y_i \in \{+1, -1\}$ with $n = 1000$. More specifically the $\{x_i\}$ are independent with $x_i \sim N(0, 1)$ with the $\{y_i\}$ generated as

$$\begin{aligned} P(y_i = +1 |\ |x_i| < 1) &= .9 \\ P(y_i = -1 |\ |x_i| > 1) &= .9 \end{aligned}$$

   (a) Implement the AdaBoost algorithm where your "weak" learners classify by simply thresholding x. That is

$$h_t(x) = \begin{cases} +1 & x < c_t \\ -1 & x > c_t \end{cases}$$

   or

$$h_t(x) = \begin{cases} +1 & x > c_t \\ -1 & x < c_t \end{cases}$$

   At each iteration t, you should choose the weak learner that gives optimal performance given the current distribution on your samples.

   (b) After 100 iterations of your algorithm plot your classifier as a function of x.

4. Construct a random data set of $K = 4$ 2-d clusters where each cluster has $n = 50$ points. The kth cluster, $k = 1, \ldots, K$ is constructed by choosing a random 2x2 matrix T_k and a random 2x1 point b_k and generating the ith example from the kth cluster, x_{ki} by

$$x_{ki} = T_k z_i + b_k$$

where the components of z_i are $N(0, 1)$. For this problem you should choose the elements of T_k to be $N(0, 1)$ and the elements of b_k to be $N(0, 10)$

   (a) Generate your points as above and plot them with different colors for each cluster.

   (b) Implement the K-means algorithm, plotting the current clusters and prototypes for each iteration. Your implmentation should pause the process to allow the user to inspect the current clustering and prototypes. This can be done by requesting character input from the user through "readline()".

   (c) Run your algorithm to convergence (no changes in clustering). Plot the number of steps to convergence using 100 random data sets.