CSE 250B: Section 5 - Sharad Vikram

1. Perceptron Practice

   Recall in the perceptron algorithm, our classifier is of the form:

$$h_w(x) = \begin{cases} +1 & \text{if } w^T x > 0 \\ -1 & \text{if } w^T x \leq 0 \end{cases}$$

   Let's try running through a few iterations of the perceptron algorithm! Let $w_0 = [0, 0]$ and recall that the update rule is if a training point (x, y) is misclassified, update $w_{t+1} = w_t + yx$. Please record the following: (a) whether or not the point is misclassified (b) the new weights, and (c) a plot of the decision boundary. The first two have been filled out for you.

   1. Training point: $x_1 = (1, 1), y_1 = +1$
      (a) Point is *misclassified*.
      (b) $w_1 = w_0 + [1, 1] = [1, 1]$
      (c) See board
   2. Training point: $x = (2, -1), y = -1$
      (a) Point is *misclassified*.
      (b) $w_2 = w_1 - [2, -1] = [-1, 2]$
      (c) See board
   3. Training point: $x = (0, 2), y = +1$

   4. Training point: $x = (-3, -1), y = -1$

   5. Training point: $x = (3, 1), y = 1$

2. Matrix calculus

   Let A be a $n \times n$ matrix, and let $x \in \mathbb{R}^n$. Show the following identities are true (the gradient will be typically be an $n \times 1$ vector):

   (a) $\nabla_x(x^T y) = y$

   (b) $\nabla_x(y^T x) = y$

   (e) $\nabla_x(x^T x) = 2x$

   (c) $\nabla_x(Ax) = A$

   (d) $\nabla_x(x^T A) = A^T$

   (f) $\nabla_x(x^T Ax) = (A + A^T)x$ (hint: use product rule)

3. Linearly Separable Data with Logistic Regression

   Show (or explain) that for a linearly separable data set, the maximum likelihood solution for the logistic regression model is obtained by finding a vector β whose decision boundary $\beta^T x = 0$ separates the classes, and taking the magnitude of β to be infinity.

4. Quadratic Kernel

   Find a feature mapping Φ such that $\Phi(x)^T \Phi(y) = K(x, y)$ where the kernel function is $K(x, y) = (x^T y + 1)^2$. For simplicity, you may assume that the data is 2-dimensional, i.e. $x = [x_1, x_2]^T$.

5. Fun with Newton's method for root-finding

   (a) Write down the iterative update equation of Newton's method for finding a root $x : f(x) = 0$ for a real-valued function f.

   (b) Prove that if $f(x)$ is a quadratic function ($f(x) = ax^2 + bx + c$), then it only takes one iteration of Newton's Method to find the minimum/maximum.