CSE 404 Hw 09

Team Evaluation:

- Brendan Rizzo
  - Completed question 1 of the homework by himself.
- **Eden Seo**
  - Completed question 2 and help Richard Huang little bit to complete question 3
- Richard Huang
  - Coded for question 3 and worked with Eden Seo to write the report.
- Jiashang Cao
  - Where is this person? I have no clue.

Questions

1. LFD example 8.13
   a. There exists a data set with two positive examples $x_1 = (0, 0)$ and $x_2 = (1, 0)$ and one negative example $x_3 = (0, 1)$. Now we look for the optimal hyperplane.
   b. The optimal solution a^* must satisfy
      i. $w^* = \sum_{n=1}^{N} y_n a_n^* x_n = y_1 a_1^* x_1 + y_2 a_2^* x_2 + y_3 a_3^* x_3 =$
      ii. $a_1 \begin{bmatrix} 0 \\ 0 \end{bmatrix} + a_2 \begin{bmatrix} 1 \\ 0 \end{bmatrix} - a_3 \begin{bmatrix} 0 \\ 1 \end{bmatrix} = [a_1 - a_3]$
   c. We can find the optimal width hyperplane to be $-2[x]_2 + 1 = 0$ with $(b, w) = (1, [0, -2])$
   d. Because $w^* = \begin{bmatrix} 0 \\ -2 \end{bmatrix}$ therefore $a_1 = 0$
   e. All three points on this hyperplane are support vectors, and you can check that for $n = 1, 2, 3,$ since we have $y_n(w^* T x_n + b^*) = 1.$
   f. Therefore, if a point (x_n, y_n) is on the boundary satisfying $y_n(w^* T x_n + b^*) = 1$ it is possible that $a_n^* = 0$ as the $a_1^* = 0$ here.

2. Part of LFD example 8.15
   a. Let us say that $\phi_1 = \begin{bmatrix} a_1 \\ \cdots \\ a_N \end{bmatrix}$, $\phi_2 = \begin{bmatrix} b_1 \\ \cdots \\ b_M \end{bmatrix}$, $K_1 = \phi_1^T \phi_1 = \sum a_i^2$, $and\ K_1 = \phi_2^T \phi_2 = \sum b_i^2$
      i. So, $\phi = \phi_1 \phi_2^T = \begin{bmatrix} a_1 \\ \cdots \\ a_N \end{bmatrix} * \begin{bmatrix} b_1 \\ \cdots \\ b_M \end{bmatrix}^T = \begin{bmatrix} a_1 b_1 & \cdots & a_1 b_M \\ \vdots & \ddots & \vdots \\ a_1 b_M & \cdots & a_N b_M \end{bmatrix}$
         1. Rewritten as $\begin{bmatrix} v_1 \\ \cdots \\ v_N \end{bmatrix}^T$ $where\ v_n = \begin{bmatrix} a_n b_1 \\ \cdots \\ a_n b_M \end{bmatrix}$
      ii. So now, we have $\phi^T \phi = \sum v_i{}^T v_i = \sum_{n=1} \sum_{m=1} a_n^2\, b_m^2 = \sum a_i^2 \sum b_i^2 = K_1 K_2$
   b. Yes. If $K_1\ and\ K_2$ are kernels, then so are $K_1 + K_2$ and $K_1 K_2$

3. Short report summarizing the observation.
    a. We used confusion matrix to calculate the accuracy of the trained model.
        i. Throughout different c values, all three svm models' accuracies are consistent, around 80% when c value is greater than 0.1. (for poly svm, degree has to be 1 to get average accuracy of 80%)
            1. When c value is lower than 0.1, all models have 56.67% accuracy rate except for linear model when c value is 0.01.
            2. The accuracy of poly svm changes as we change the degree of it. As we increase the degree, the lower the accuracy it gets.
                a. If degree is 5, the average accuracy is 73%
                b. If degree is 10, the average accuracy is 68%
                c. If degree is 100, the average accuracy is 55%
    b. When c value is above 1, linear svm model and poly svm model have approximately the same amount of support vectors, around 50, but rbf svm model (default kernel) has more support vectors, around 70. (for poly svm, degree has to be 1 in order to get this value)
        i. When c value is lower than 1
            1. When c value is between 0.0001 and 0.001, they all have the same amount of support vector, around 136.
            2. When c value is 0.01, linear svm model has 107 support vector and rest of the two have 137 support vectors.
            3. When c value is 0.1, linear svm model has 64, rbf svm model has 123, and poly support vector has 98
        ii. When degree is bigger than 1 for poly svm model, amount of support vector changes.
            1. When degree is 5, average amount of support vector is 80
            2. When degree is 10, average amount of support vector is 100
            3. When degree is 10, average amount of support vector is 115