Kyler Little
CPT_S 437

Homework #1

1.) Consider the perceptron in two dimensions: $h(x) = \text{sign}(w^T x)$ where $w = [w_0, w_1, w_2]^T$ and $x = [1, x_1, x_2]^T$. Technically, x has three coordinates, but we call this perceptron two-dimensional because the first coordinate is fixed at 1.

(a) Show that the regions on the plane where $h(x) = +1$ and $h(x) = -1$ are separated by a line. If we express this line by the equation $x_2 = ax_1 + b$, what are the slope a and intercept b in terms of w_0, w_1, w_2?

Consider that the 'sign' function returns +1 when the argument provided is greater than 0, and it returns -1 when the argument provided is less than 0. Thus, when the argument is 0, the domain is divided into two halves. These halves are called "half spaces." Half spaces are created by an $n - 1$ dimensional hyperplane. In this case, the hyperplane is a simple line in R^2. That line is:

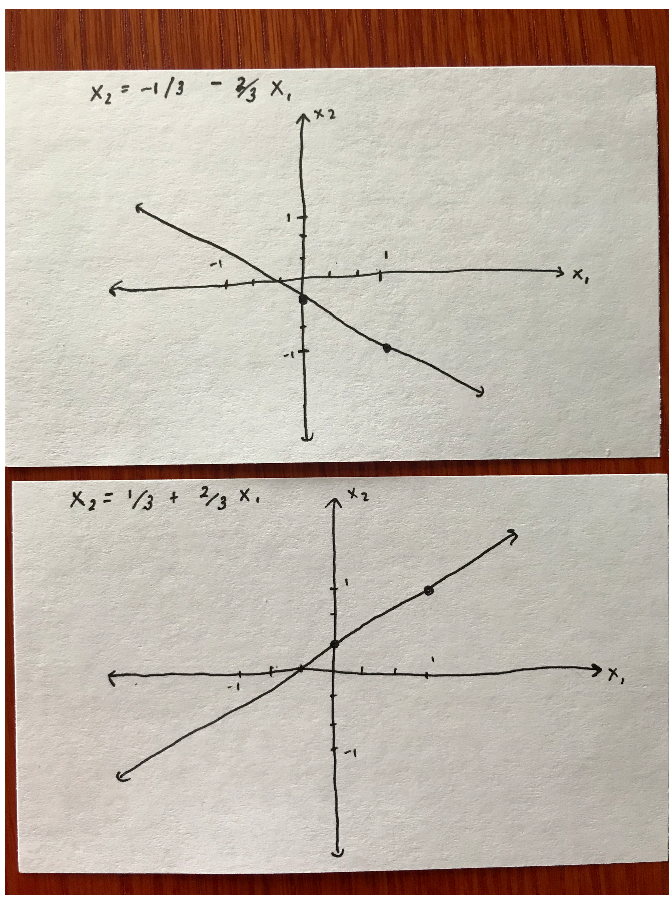
$$0 = w^T x = w_0 + w_1 x_1 + w_2 x_2$$

Solving for x_2 yields:

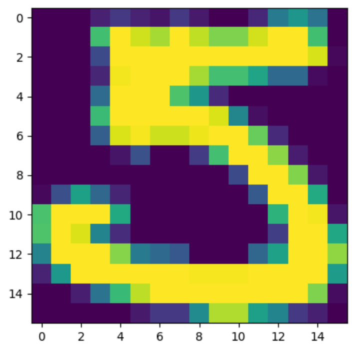
$$x_2 = -\frac{w_0 + w_1 x_1}{w_2}$$

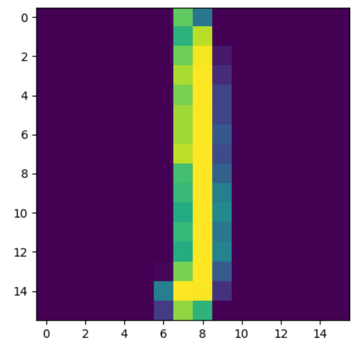
Thus, $a = -w_1/w_2$ and $b = -w_0/w_2$.

(b) Draw a picture for the cases $w = [1,2,3]^T$ and $w = -[1,2,3]^T$.

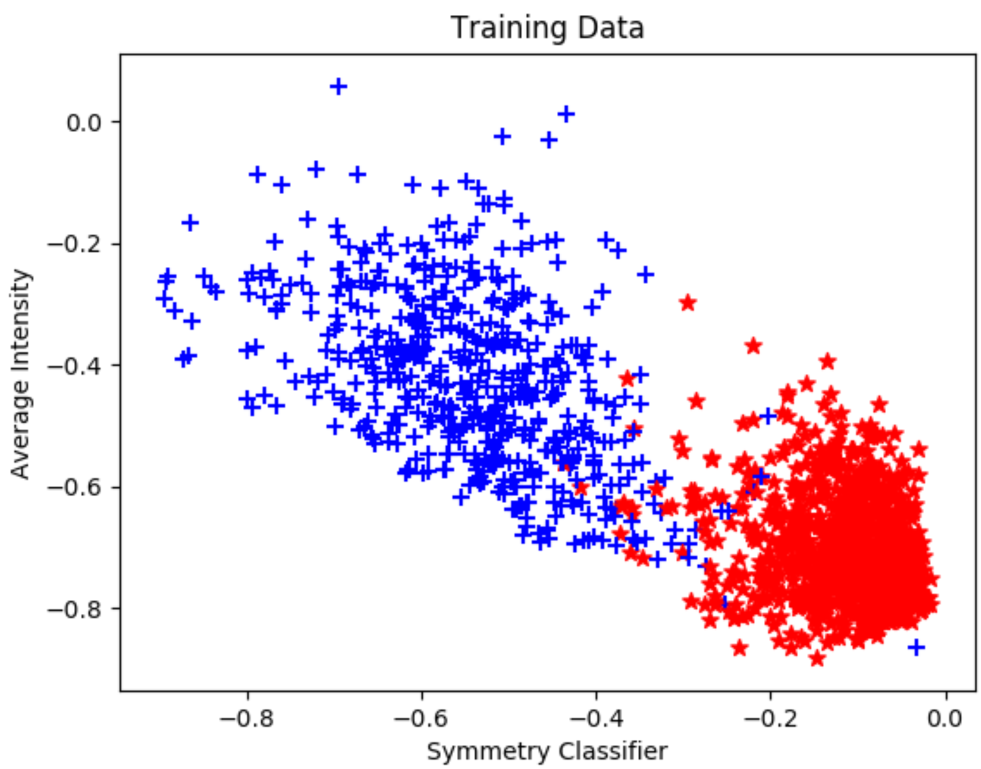


2.) In this problem, we use a perceptron learning model to differentiate between the handwritten digits of '1' and '5'. Here are two examples of what the digits actually look like.



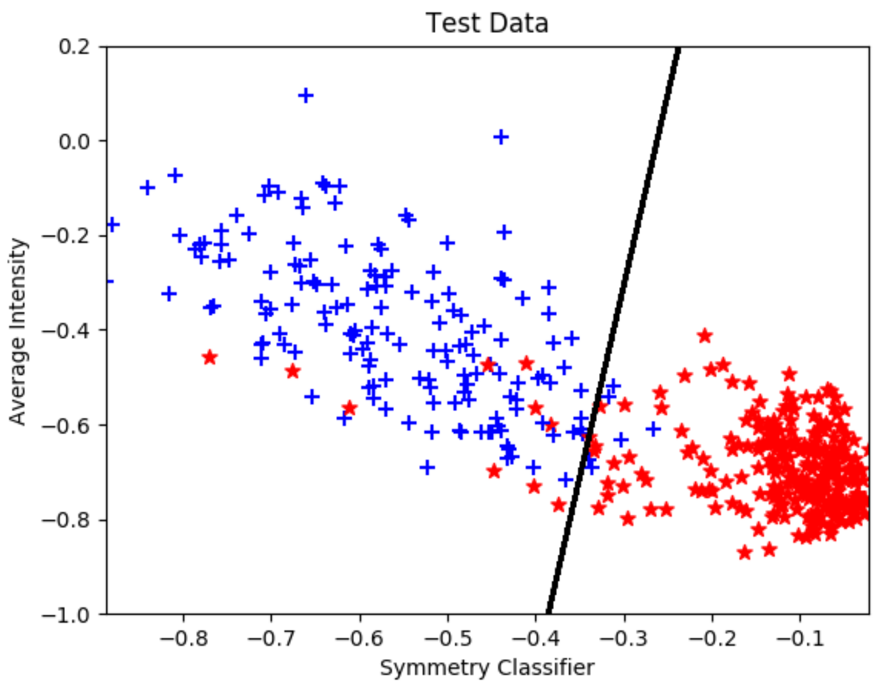


To distinguish between the two, we used about 1500 labeled samples to train our model. For each image, we classify the digit by its symmetry—since a '1' and '5' have very different symmetries—and its average intensity—since it takes more pen ink to write a '5' than it does for a '1'. In general, if these features make the data linearly separable, then the feature choices were good. In this case, it is easy to see that our choices made the training data approximately linearly separable.



Above, the blue '+' symbols correspond with the '5' digit; the red '*' symbols correspond with the '1' digit.

To test if our model is correct (i.e. it classifies unknown data samples mostly correctly), we used about 400 more samples. These samples were also already labeled, so in order to test our model, we compared the correct labels to our model's predicted labels. In the graph below, the black line represents our model. A blue '+' symbol to the left of the line indicates the '5' digit was correctly classified; a red '*' symbols to the right of the line indicates the '1' digit was correctly classified. Otherwise, the digit was misclassified.



It is easy to see that our model does a good job for the majority of test samples.

3.) Given $x \in R^m, y \in R^n$, show that the rank of matrix xy^T is one.
Let's actually expand out the multiplication to see what it looks like.

$$xy^T = \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_m \end{bmatrix} [y_1 \quad y_2 \quad \dots \quad y_n] = \begin{bmatrix} x_1y_1 & x_1y_2 & \dots & x_1y_n \\ x_2y_1 & x_2y_2 & \dots & \dots \\ \dots & \dots & \dots & \dots \\ x_my_1 & \dots & \dots & x_my_n \end{bmatrix}$$

From this, it is easy to see the linear dependence. Notice how y^T is present in each column of the matrix and how x is present in each column. In fact, the only difference between any given column or row in the matrix is the factor that scales it. Thus, it is easy to write any row/column as a linear combination of other rows/columns. For instance, let $\alpha = x_1/x_2$ and notice that $x_1y^T = \alpha x_2 y^T$ immediately; the first row is a linear combination of the second row. In general,

$$x_i y^T = \alpha x_j y^T$$

where $\alpha = x_i/x_j$, $i, j \in \{1,2,\dots,m\}$, and $i \neq j$, effectively making each row a linear combination of the others. Likewise,

$$y_i x = \beta y_j x$$

where $\beta = y_i/y_j$, $i, j \in \{1,2,\dots,n\}$, and $i \neq j$.
Therefore, the maximum number of linearly independent columns/rows of xy^T is one, so the rank is one.

4.) Given $X = [x_1, x_2, \dots, x_n] \in R^{mxn}$ where $x_i \in R^m$ for all i, and $Y^T = [y^1, y^2, \dots, y^n] \in R^{pxn}$ where $y^i \in R^p$ for all i. Show that

$$XY = \sum_{i=1}^{n} x_i \left(y^i\right)^T$$

To start, I will expand the product XY. In the following notation, let x_j^i denote the element in the jth column and ith row.

$$XY = \begin{bmatrix} x_1^1 & x_2^1 & \dots & x_n^1 \\ x_1^2 & x_2^2 & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \dots & \dots & x_{n-1}^{m-1} & x_n^{m-1} \\ x_1^m & \dots & x_{n-1}^m & x_n^m \end{bmatrix} \begin{bmatrix} y_1^1 & y_2^1 & \dots & \dots & y_p^1 \\ y_1^2 & y_2^2 & \dots & \dots & \dots \\ \dots & \dots & \dots & y_{p-1}^{n-1} & y_p^{n-1} \\ y_1^n & \dots & \dots & y_{p-1}^n & y_p^n \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} x_i^1 y_1^i & \sum_{i=1}^{n} x_i^1 y_2^i & \dots & \sum_{i=1}^{n} x_i^1 y_p^i \\ \sum_{i=1}^{n} x_i^2 y_1^i & \sum_{i=1}^{n} x_i^2 y_2^i & \dots & \dots \\ \dots & \dots & \dots & \sum_{i=1}^{n} x_i^{m-1} y_p^i \\ \sum_{i=1}^{n} x_i^m y_1^i & \dots & \sum_{i=1}^{n} x_i^m y_{p-1}^i & \sum_{i=1}^{n} x_i^m y_p^i \end{bmatrix}$$

Since each sum indexes the same, we can factor the summation out of the matrix. Then, we can remove the changing indices by noting that

$$\begin{bmatrix} x_i^1 y_1^i & x_i^1 y_2^i & \dots & x_i^1 y_p^i \end{bmatrix} = x_i^1 \left(y^i\right)^T$$

and

$$\begin{bmatrix} x_i^1 y_1^i \\ x_i^2 y_1^i \\ \dots \\ x_i^m y_1^i \end{bmatrix} = x_i y_1^i$$

Thus, we have:

$$XY = \sum_{i=1}^{n} x_i \left(y^i\right)^T$$

5.) Given $X \in R^{mxn}$, show that the matrix $X^T X$ is symmetric and positive semi definite. When is it positive definite?

Let $X' = X^T X$.

$$X' = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{m1} \\ x_{12} & \dots & \dots & \dots \\ \dots & \dots & x_{(m-1)(n-1)} & x_{m(n-1)} \\ x_{1n} & \dots & x_{(m-1)n} & x_{mn} \end{bmatrix} \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & \dots & \dots & \dots \\ \dots & \dots & x_{(m-1)(n-1)} & x_{(m-1)n} \\ x_{m1} & \dots & x_{m(n-1)} & x_{mn} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{m} x_{i1} x_{i1} & \sum_{i=1}^{m} x_{i1} x_{i2} & \dots & \sum_{i=1}^{m} x_{i1} x_{in} \\ \sum_{i=1}^{m} x_{i2} x_{i1} & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \sum_{i=1}^{m} x_{in} x_{i1} & \dots & \dots & \sum_{i=1}^{m} x_{in} x_{im} \end{bmatrix}$$

Clearly, the matrix is symmetric since multiplication is commutative. To prove the matrix is positive semi definite, we must show that $x^T X' x \geq 0, \forall x \in R^m$.

Recall that we defined X' as $X^T X$. Note that:

$$x^T X^T X x = (Xx)^T X x = \left|\left|Xx\right|\right|_2$$

Since the L_2 is always greater than or equal to 0, we have that:

$$x^T X^T X x = x^T X' x \geq 0$$

Thus X' is positive semi definite. If $Xx \neq 0$, then

$$x^T X^T X x = x^T X' x > 0$$

and X' would be positive definite. Furthermore, because $x \neq 0$ to be positive definite, this means that $X \neq 0 \in R^{mxn}$ is a sufficient condition for X' to be positive definite.

6.) Given $g(x, y) = e^x + e^{y^2} + e^{3xy}$, compute $\frac{\partial g}{\partial y}$.

$$\frac{\partial g}{\partial y} = 2ye^{y^2} + 3xe^{3xy}$$

7.) Consider the matrix A.

$$A = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 1 & 2 \\ 3 & 2 & 5 \end{bmatrix}$$

(a) Compute the eigenvalues and corresponding eigenvectors of A. You are allowed to use Matlab to compute the eigenvectors (but not the eigenvalues).

To find the eigenvalues, we first must find the roots of the characteristic polynomial:

$$|A - \lambda I| = 0$$

$$\begin{aligned}
|A - \lambda I| &= \begin{vmatrix} 2-\lambda & 1 & 3 \\ 1 & 1-\lambda & 2 \\ 3 & 2 & 5-\lambda \end{vmatrix} \\
&= (2-\lambda)\begin{vmatrix} 1-\lambda & 2 \\ 2 & 5-\lambda \end{vmatrix} - \begin{vmatrix} 1 & 2 \\ 3 & 5-\lambda \end{vmatrix} + 3\begin{vmatrix} 1 & 1-\lambda \\ 3 & 2 \end{vmatrix} \\
&= (2-\lambda)\big((1-\lambda)(5-\lambda) - 4\big) - \big((5-\lambda) - 6\big) + 3\big(2 - 3(1-\lambda)\big) \\
&= -\lambda^3 + 8\lambda^2 - 3\lambda \\
&= 0
\end{aligned}$$

Finding the roots is a simple matter of the quadratic formula. Factor out a $-\lambda$ to yield the first eigenvalue: $\lambda_1 = 0$. Then, use the quadratic formula on the remainder of the characteristic polynomial:

$$0 = -\lambda(\lambda^2 - 8\lambda + 3)$$

This yields $\lambda_2 = -\sqrt{13} + 4$ and $\lambda_3 = \sqrt{13} + 4$.

To find the eigenvectors, I used Matlab.

$$v_1 = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}, v_2 = \begin{bmatrix} -\sqrt{13} - 3 \\ \sqrt{13} + 4 \\ 1 \end{bmatrix}, v_3 = \begin{bmatrix} \sqrt{13} - 3 \\ -\sqrt{13} + 4 \\ 1 \end{bmatrix}$$

(b) What is the eigen-decomposition of A?

I used Matlab to find the inverse of the matrix.

$$A = \begin{bmatrix} -1 & -\sqrt{13} - 3 & \sqrt{13} - 3 \\ -1 & \sqrt{13} + 4 & -\sqrt{13} + 4 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & -\sqrt{13} + 4 & 0 \\ 0 & 0 & \sqrt{13} + 4 \end{bmatrix} \begin{bmatrix} -1/3 & -1/3 & 1/3 \\ \frac{13 - 5\sqrt{13}}{78} & \frac{13 - 2\sqrt{13}}{78} & \frac{26 - 7\sqrt{13}}{78} \\ \frac{13 + 5\sqrt{13}}{78} & \frac{2 + \sqrt{13}}{6\sqrt{13}} & \frac{26 + 7\sqrt{13}}{78} \end{bmatrix}$$

(c) What is the rank of A?

To find the rank of A, we row-reduce A to its upper echelon form and count the number of pivot columns.

$$A = \begin{bmatrix} 2 & 1 & 3 \\ 1 & 1 & 2 \\ 3 & 2 & 5 \end{bmatrix}$$

$$A = \begin{bmatrix} 1 & 1/2 & 3/2 \\ 1 & 1 & 2 \\ 3 & 2 & 5 \end{bmatrix}, r_1 = r_1/2$$

$$A = \begin{bmatrix} 1 & 1/2 & 3/2 \\ 0 & 1 & 1 \\ 3 & 2 & 5 \end{bmatrix}, r_2 = 2(r_2 - r_1)$$

$$A = \begin{bmatrix} 1 & 1/2 & 3/2 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} r_3 = 2(r_3 - 3r_1) - r_2$$

There are two pivot columns, so the rank of A is two.

(d) Is A positive definite? Is A positive semi-definite?

For A to be positive semi-definite, we must have that $x^T A x \geq 0, \forall x \in R^3$. Let's pick an arbitrary vector $x = [a \quad b \quad c]^T$.

$$\begin{aligned} x^T A x &= [a \quad b \quad c]^T \begin{bmatrix} 2 & 1 & 3 \\ 1 & 1 & 2 \\ 3 & 2 & 5 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} \\ &= [2a + b + 3c \quad a + b + 2c \quad 3a + 2b + 5c] \begin{bmatrix} a \\ b \\ c \end{bmatrix} \\ &= 2a^2 + b^2 + 3c^2 + a^2 + b^2 + 2c^2 + 3a^2 + 2b^2 + 5c^2 \\ &= 6a^2 + 4b^2 + 10c^2 \end{aligned}$$

Since this is just a sum of squares, we know that $x^T A x \geq 0$. In fact, $x^T A x > 0$ when $x \neq 0$. Thus, A is both positive definite and positive semi-definite.

(e) Is A singular?

A is not nonsingular since $\text{rank}(A) \neq n = 3$; therefore, A is singular.