EECS 738/800: Homework 4

2017-04-12

Name: Elise McEllhiney Collaborators: David Young, David Ménager, Bekah Coggins

1. Consider the data below and a 'hyperplane' $(b, \mathbf{w})$ that separates the data.

$$X = \begin{bmatrix} 0 & 0 \\ 2 & 2 \\ 2 & 0 \end{bmatrix} \qquad y = \begin{bmatrix} -1 \\ -1 \\ +1 \end{bmatrix} \qquad \mathbf{w} = \begin{bmatrix} 1.2 \\ -3.2 \end{bmatrix} \qquad b = -0.5$$

   (a) Compute $\rho = \min_{i=1,\dots,N} y_i(\mathbf{w}^T\mathbf{x}_i + b)$.

$$\rho = \min_{i=1,\dots,N} \begin{bmatrix} -1 \\ -1 \\ +1 \end{bmatrix} \left(\begin{bmatrix} 1.2 & -3.2 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 2 & 2 \\ 2 & 0 \end{bmatrix} + (-0.5) \right) \tag{1}$$

$$= \min_{i=1,\dots,N} \begin{bmatrix} -1 \\ -1 \\ +1 \end{bmatrix} \left(\begin{bmatrix} 0 \\ -4 \\ 2.4 \end{bmatrix} + (-0.5) \right) \tag{2}$$

$$= \min_{i=1,\dots,N} \begin{bmatrix} -1 \\ -1 \\ +1 \end{bmatrix} \begin{bmatrix} -0.5 \\ -4.5 \\ 1.9 \end{bmatrix} \tag{3}$$

$$= \min_{i=1,\dots,N} \begin{bmatrix} 0.5 \\ 4.5 \\ 1.9 \end{bmatrix} \tag{4}$$

$$\rho = 0.5 \tag{5}$$

   (b) Compute the weights $\frac{1}{\rho}(b, \mathbf{w})$ and show that they satisfy $\min_{i=1,\dots,N} y_i(\mathbf{w}^T\mathbf{x}_i + b) = 1$.

$$\min_{i=1,\dots,N} \begin{bmatrix} -1 \\ -1 \\ +1 \end{bmatrix} \left(\begin{bmatrix} 2.4 & -6.4 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 2 & 2 \\ 2 & 0 \end{bmatrix} + (-1) \right) = \min_{i=1,\dots,N} \begin{bmatrix} -1 \\ -1 \\ +1 \end{bmatrix} \left(\begin{bmatrix} 0 \\ -8 \\ 4.8 \end{bmatrix} + (-1) \right) \tag{6}$$

$$= \min_{i=1,\dots,N} \begin{bmatrix} -1 \\ -1 \\ +1 \end{bmatrix} \begin{bmatrix} -1 \\ -9 \\ 3.8 \end{bmatrix} \tag{7}$$

$$= \min_{i=1,\dots,N} \begin{bmatrix} 1 \\ 9 \\ 3.8 \end{bmatrix} \tag{8}$$

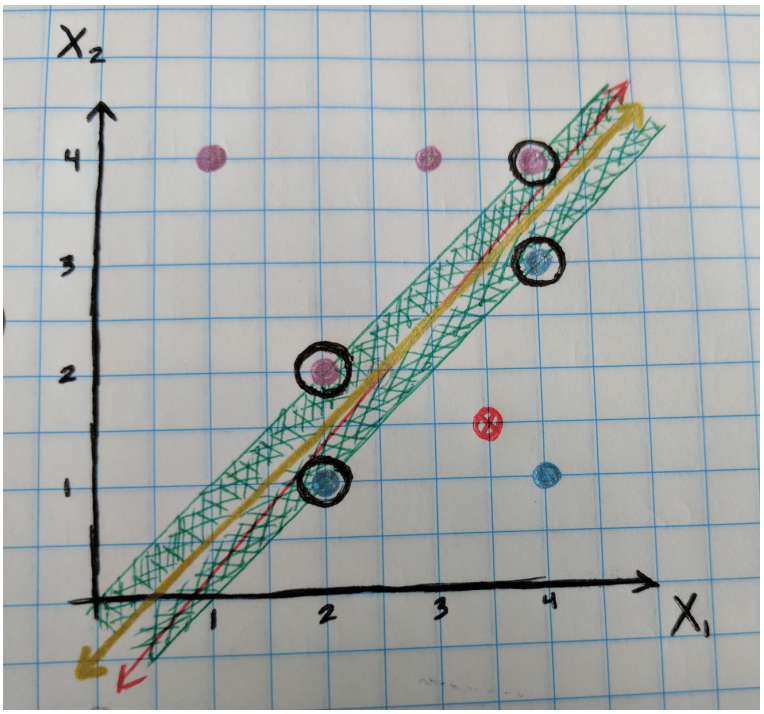
$$= 1 \tag{9}$$

2. Here we explore the maximal margin classifier (linearly separable SVM) on a toy data set.

   (a) We are given $n = 7$ observations in $p = 2$ dimensions. For each observation, there is an associated class label.

Obs	X_1	X_2	y
1	3	4	Red
2	2	2	Red
3	4	4	Red
4	1	4	Red
5	2	1	Blue
6	4	3	Blue
7	4	1	Blue

Sketch the observations.



(b) Sketch the optimal separating hyperplane and provide the equation for this hyperplane of the form $b + w_1X_1 + w_2X_2 = 0$.

$$b + w_1X_1 + w_2X_2 = 0 \tag{10}$$
$$-0.5 + X_1 - X_2 = 0 \tag{11}$$

(c) Describe the classification rule for maximal margin (linear SVM) classifier. It should be something like "Classify to Red if $b + w_1X_1 + w_2X_2 \geq 0$ and classify to Blue otherwise. Provide $\mathbf{w}$ and b.

$$b = -0.5 \tag{12}$$
$$\boldsymbol{w} = [1, -1] \tag{13}$$

Classify to Red if $-0.5 + X_1 - X_2 \leq 0$ and to Blue otherwise.

(d) On your sketch from the first part, indicate the margin for the hyperplane.
The margin is the green region

(e) Indicate the support vectors for the classifier.
The support vectors are the circled points.

(f) Argue that a slight movement of the seventh observation would or would not affect the hyperplane of our linear SVM.
Any point besides the support vectors will not effect the hyperplane by moving unless the movement causes the point to enter the margin or cross the hyperplane. The point entering the margin would change the ideal hyperplane and reduce our margin since the point would become a support vector by being that close to the separating hyperplane. Crossing the hyperplane would cause the point to be misclassified. So for the 7th observation, it would have to move more than 3 units in one or both axis before it would even touch the margin. So any deviation less than this wouldn't affect the hyperplane.

(g) Sketch a hyperplane that separates the data but is not the optimal hyperplane. Provide an equation for this hyperplane.

The thin bright red line is a non-optimal hyperplane.

$$-5 + 6X_1 - 5X_2 = 0$$

(h) Draw an additional observation on the plot so that the two classes are no longer separable by a hyperplane. Clearly label that point.

The bright red XOR symbol, if classified as "Red" would prevent the observations from being separable

3. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) can both be used for dimensionality reduction.

(a) Explain what the first principal component from PCA captures.

The first component from PCA is the eigenvector that captures the greatest variance.

(b) How does PCA differ from LDA?

PCA determines component axes that maximize the variance. These new axis are then used for class separation. LDA, on the other hand, only maximizes the existing component axes for class separation.

(c) Give a short description of why PCA can be used for image compression without much loss of image quality.

In image compression, PCA can be used without much loss in quality since past a reasonable number of principal components, the variance explained by additional components would not produce a notable change in the image.

4. Neural networks with a single hidden unit and a logistic activation function have real-valued inputs $X_1, \ldots, X_n$ where the unit output Y is defined as

$$Y = \frac{1}{1 + \exp(-(w_0 + \sum_i w_i X_i))}.$$

Let the X_i be either 0 or 1. We will assume that a value for $Y > 0.5$ is 1 and zero otherwise. So even though our neural network returns a probability, we will convert the probability to a 0 or 1.

(a) Give 3 weights for a single unit with two inputs X_1 and X_2, that implements the logical "OR" function $Y = X_1 \vee X_2$.

Yes, this can be implemented in a single unit.

$$\boldsymbol{w} = [-1, 2, 2]$$

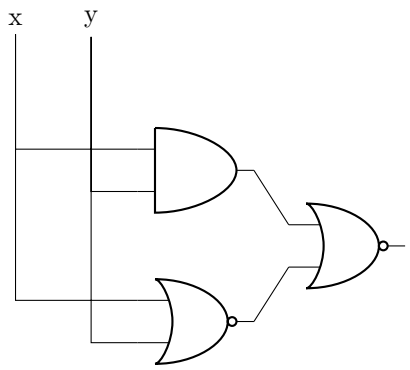
(b) Can you implement the logical "AND" function $Y = X_1 \wedge X_2$ with a single hidden unit? If so give weights that achieve this. If not, explain the problem.

Yes, this can be implemented in a single unit.

$$\boldsymbol{w} = [-3, 2, 2]$$

(c) Can you implement the logical "EXCLUSIVE-OR" function $Y = X_1 \bigoplus X_2$ with a single hidden unit? If so give weights that achieve this. If not, how many hidden units would you need?

No, this cannot be implemented with a single unit. You would need two hidden units. I assume that outputs of the neural units are 1 or 0.



$$\text{Hidden unit 1:} \quad \boldsymbol{w} = [-3, 2, 2] \tag{14}$$

$$\text{HIdden unit 2:} \quad \boldsymbol{w} = [1, -2, -2] \tag{15}$$

$$\text{Output node:} \quad \boldsymbol{w} = [1, -2, -2] \tag{16}$$