Stamatios Morellas

COM S 474 - Midterm Exam

10/18/2020

Pledge

> "I affirm that the work on this exam is my own and I will not use any people to help me nor will I share any part of this exam or my work with others without permission of the instructor."

Question 1

The three types of machine learning are ***Supervised***, ***Unsupervised***, and ***Reinforcement*** Learning.

Question 2

To find the class for the linear binary classifier $\mathbf{w}^T x > 0$, we must plug in the values for the weight and feature vectors. Doing this, we get the following:

$$\begin{pmatrix} 1 & 2 & 3 \end{pmatrix} * \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = 1 + 2 + 3 = \mathbf{6}$$

Since $6 > 0$, this falls into the class $+\mathbf{1}$.

Question 3

To find the error of the classifier on sample x, we use the equation:

$$J(\mathbf{W}) = \sum_{i=0}^{N} (\mathbf{w}^T x_i - y_i)^2$$

Plugging in the values for $\mathbf{w}^T$ and x, we get the following:

$$(\begin{pmatrix} 1 & 1 & 1 \end{pmatrix} * \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} - (-1))^2$$

Simplifying this, we get:

$$(6 + 1)^2 = \mathbf{49}$$

Question 4

This loss function: $\sum(\hat{y} - y)^2$ treats all the misclassified samples equally by contributing to a total of 4 for every misclassified sample. However, this loss function: $\sum(\mathbf{w}^T x - y)^2$ weighs more misclassified samples that are away from the decision surface.

Question 5

For the *first* case, $Pr(class = +1|b > 5)$, the samples that pass the condition $C1 : b > 5$ are samples S_3 and S_6, where S_i denotes the i^{th} sample. Given this, we now get:

$$Pr(class = +1|b > 5) = Pr(class = +1|S_3 = +1, S_6 = -1) = \mathbf{1/2}$$

For the *second* case, $Pr(class = -1|b > 5)$, the samples that pass the condition $C2 : b > 5$ are samples S_3 and S_6, where S_i denotes the i^{th} sample. Given this, we now get:

$$Pr(class = -1|b > 5) = Pr(class = -1|S_3 = +1, S_6 = -1) = \mathbf{1/2}$$

For the *third* case, $Pr(class = +1|b <= 5)$, the samples that pass the condition $C3 : b <= 5$ are samples S_1, S_2, S_4, and S_5, where S_i denotes the i^{th} sample. Given this, we now get:

$$Pr(class = +1|b <= 5) = Pr(class = +1|S_1 = +1, S_2 = +1, S_4 = -1, S_5 = -1) = 2/4 = \mathbf{1/2}$$

Finally, for the *fourth* case, $Pr(class = -1|b <= 5)$, the samples that pass the condition $C4 : b <= 5$ are samples S_1, S_2, S_4, and S_5, where S_i denotes the i^{th} sample. Given this, we now get:

$$Pr(class = -1|b <= 5) = Pr(class = -1|S_1 = +1, S_2 = +1, S_4 = -1, S_5 = -1) = 2/4 = \mathbf{1/2}$$

We will now compute the Gini impurity for feature b with threshold 5.

For the *left* group, the samples that pass the condition $b <= 5$ are

$$S_1 = +1, S_2 = +1, S_4 = -1, S_5 = -1$$

For the *right* group, the samples that pass the condition $b > 5$ are

$$S_3 = +1, S_6 = -1$$

The Gini impurity for the left and right groups are as follows:

$$G_{left} = 2/4 * (1 - 2/4) + 2/4 * (1 - 2/4) = 1/2 = 0.5$$

$$G_{right} = 1/2 * (1 - 1/2) + 1/2 * (1 - 1/2) = 1/2 = 0.5$$

Question 6

The probability that $b > 5$, i.e. $Pr(b > 5) = 2/6 = 1/3$

The probability that $b <= 5$, i.e. $Pr(b <= 5) = 4/6 = 2/3$

Question 7

The expectation for the total Gini impurity is given by the following:

$$G_{total} = 4/6 * G_{left} + 2/6 * G_{right} = (4/6 * 1/2) + (2/6 * 1/2) = \mathbf{1/2}$$

Question 8

The samples that will be chosen to become the support vectors are the ones that satisfy the equation $\lambda_i \neq 0$, which means that λ_1, λ_3, and λ_4 will be chosen as the support vectors.

Question 9

We use the following equation to determine the properties of a hard margin linear SVM:

$$\mathbf{w} = \sum_{x_k \in N_s} \lambda_k y_k \mathbf{x_k}$$

Given this, we get $\mathbf{w} = \lambda_1 y_1 x_1 + \lambda_3 y_3 x_3 + \lambda_4 y_4 x_4$ since $\lambda_2 = 0$.

Plugging in the respective values, we get:

$$\mathbf{w} = (6.13*(+1)*\begin{pmatrix} 0.5 \\ 0.25 \\ 0.125 \end{pmatrix})+(4.08*(-1)*\begin{pmatrix} 0.3 \\ 0.75 \\ 0.325 \end{pmatrix})+(2.05*(-1)*\begin{pmatrix} 0.2 \\ 0.65 \\ 0.425 \end{pmatrix}) = \begin{pmatrix} 1.431 \\ -2.86 \\ -1.431 \end{pmatrix}$$

Question 10

To compute the predictions, we compute $\mathbf{w^T x_i} + w_b$ for each sample above.

For sample S_1:

$$\mathbf{w^T x_i} + w_b = \begin{pmatrix} 1.431 & -2.86 & -1.431 \end{pmatrix} \begin{pmatrix} 0.5 \\ 0.25 \\ 0.125 \end{pmatrix} + 1.18 = \mathbf{1.0015}$$

For sample S_2:

$$\mathbf{w^T x_i} + w_b = \begin{pmatrix} 1.431 & -2.86 & -1.431 \end{pmatrix} \begin{pmatrix} 0.4 \\ 0.15 \\ 0.225 \end{pmatrix} + 1.18 = \mathbf{1.0012}$$

For sample S_3:

$$\mathbf{w^T x_i} + w_b = \begin{pmatrix} 1.431 & -2.86 & -1.431 \end{pmatrix} \begin{pmatrix} 0.3 \\ 0.75 \\ 0.325 \end{pmatrix} + 1.18 = -\mathbf{1.0011}$$

For sample S_4:

$$\mathbf{w^T x_i} + w_b = \begin{pmatrix} 1.431 & -2.86 & -1.431 \end{pmatrix} \begin{pmatrix} 0.2 \\ 0.65 \\ 0.425 \end{pmatrix} + 1.18 = -\mathbf{1.0014}$$

Question 11

In order for a sample to fall into the margin a.k.a "gutter", the following condition must hold:

$$-1 <= \mathbf{w^T x_i} + w_b <= 1$$

Given that $S_1 = 1.0015$, $S_2 = 1.0012$, $S_3 = -1.0011$, and $S_4 = -1.0014$, this means that ***none*** of the samples fall into the margin, although they are very close to the margin.

Question 12

One thing that seemed odd is that each of the samples provided were correctly classified, however, they all appeared to have the same approximate distance of 0.001 from the gutter lines.