Question 10.

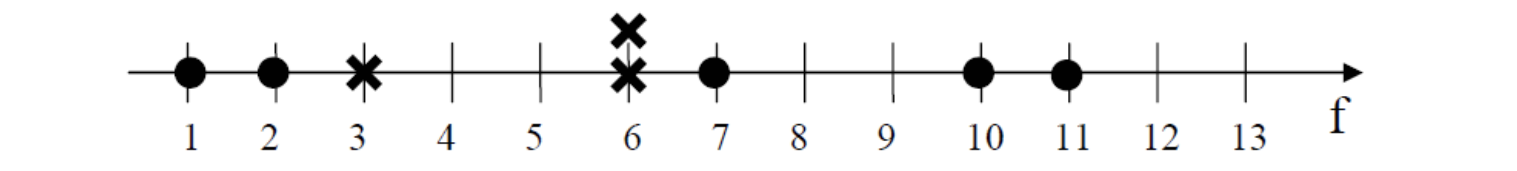
Question 10

The diagram below shows a dataset with 2 classes and 8 data points, each with only one feature

value, labeled f. Note that there are two data points with the same feature value of 6. These are

shown as two x’s one above the other. Provide stepwise mathematical solution, do not write

code for it.



1. Divide this data equally into two parts. Use first part as training and second part as

testing. Using KNN classifier, for K=3, what would be the predicted outputs for the test

samples? Show how you arrived at your answer.

2. Compute the confusion matrix for this and calculate accuracy, sensitivity and specificity

values.

**Answer**

**Given:**

Total classes = 2, Considering the two classes as class A and Class B

Data points =8

Feature = 1 f

K=3

**Explanation for the predicted outputs for test samples:**

**Step 1:** K= 3 which means, considering the test sample has 3 nearest neighbors.

We have 2 datapoints on same feature value.

**Step 2:** Dividing the dataset into Training data and test data.

Given Training data is TR = 1,2,3,4,5,6,7,8,9,10,11,12,13

And I have considered Test Samples as TS = 1, 3, 5, 7, 9, 11

Consider, Class A is

Class B is

According to the data model, the training data is classified as below.

Training Data (TD)

|  |  |
| --- | --- |
| Input | Target (Class) |
| 1 | A |
| 2 | A |
| 3 | B |
| 6 | B |
| 6 | B |
| 7 | A |
| 10 | A |
| 11 | A |

According to the training data, the test data samples, outputs can be predicted as below.

For TS=1, Neighbors are TD=1, TD=2, TD=3-> Output Class = A

For TS=3, Neighbors are TD=2, TD=3, TD=4, -> Output Class = B

For TS=5, Neighbors are TD=4, TD=5, TD=6, TD = -> Output Class = B

For TS=7, Neighbors are TD=7, TD=8 -> Output Class = B

For TS=9, Neighbors are TD=9, TD=10 -> Output Class = A

For TS=11, Neighbors are TD=12, TD=13 -> Output Class = A

So, the below table will be the outputs of test data.

|  |  |
| --- | --- |
| Test Data Input | Predicted output class |
| 1 | A |
| 3 | B |
| 5 | B |
| 7 | B |
| 9 | A |
| 11 | A |

Finally, the predicted outputs of test data samples can be both A and B class.

2. We can calculate confusion matrix by comparing predicted outputs with actual output using below formula.

Based on above values we can compute Confusion Matrix as -

True Positives (TP) = 1(Output of 3 in Test Set is B & Predicted Output by KNN is also B)

True Negatives (TN)= 0 (Output of 1 in Test Set is A & Predicted Output by KNN is also A) False Positives (FP) = 0 (Output of 7in Test Set is A & Predicted Output by KNN is B) False Negatives (FN)= 1 (Output in Test Set is B & Predicted Output by KNN is A)

**Confusion Matrix:** Confusion Matrix is the matrix as output and describes the complete performance of the model.

|  |  |  |
| --- | --- | --- |
|  | **A (0)** | **B (1)** |
| **0** | 0(TN) | 0(FN) |
| **1** | 0(FP) | 1(TP) |

**Accuracy Rate =** (Number of Correct Predictions / Total Number of Predictions) × 100%

= (TP+TN/P+N) ×100%

= (1+01+0+0+1+1)/8 ×100%

=50%

**Sensitivity or True Positive Rate or Recall Rate:** True Positive Rate is defined as*TP/ (FN+TP)*. True Positive Rate corresponds to the proportion of positive data points that are correctly considered as positive, with respect to all positive data points.

=TP/TP+FN

=1/1+1

=1/2

=50%

**Specificity or True Negative Rate:** True Negative Rate is defined as *TN / (FP+TN)*. False Positive Rate corresponds to the proportion of negative data points that are correctly considered as negative, with respect to all negative data points.

=(TN/TN+FP) x100

= (0/0+0)×100

=Undefined