Course: Machine Learning - Theory
Week 4 Graded assignment solution

1. (1 point) Which of the following is/are the evaluation measures for classification models?

   A. Confusion matrix.

   B. Precision, recall, F1 scores, accuracy.

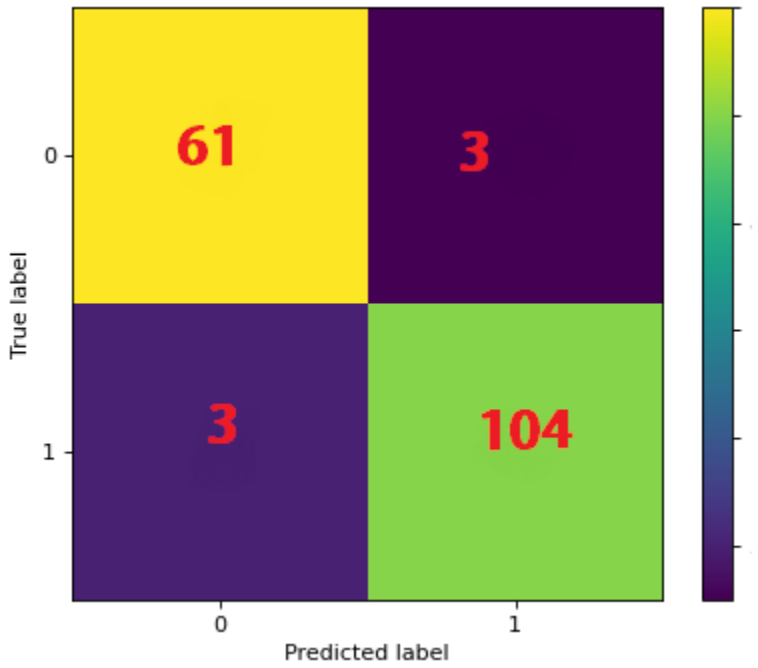
   C. AUC curve.

   D. All of the above.

Answer: D

Solution:
All of the above metrics are useful to evaluate classification models.

(Common data for Q2 ,Q3,Q4) To evaluate the performance of a trained model confusion matrix is plotted, which is given below.



2. (1 point) Based on the above plotted confusion matrix what is the precision score of the trained model?

A. 0.972

B. 0.965

C. 0.5865

D. 0.598

Answer: A

Solution: Precision= $\frac{TP}{TP+FP}$ =$\frac{104}{104+3}$=0.9719

3. (1 point) Based on the above plotted confusion matrix what is the accuracy score of the trained model?

A. 0.972

B. 0.965

C. 0.5865

D. 0.598

Answer: B

Solution: Accuracy= $\frac{TP+TN}{TP+FP+TN+FN}$=$\frac{104+61}{104+3+61+3}$=0.9649

4. (1 point) Based on the above plotted confusion matrix what is the F1 score of the trained model?

A. 0.972

B. 0.965

C. 0.5865

D. 0.598

Answer: A

Solution: F1 score= $\frac{2*precision*recall}{precision+recall}$

5. (1 point) In Multi-label classification problems each instance can be assigned multiple categories or a set of target labels.

A. True

B. False

Answer: True

Solution: Yes in Multilabel classification each instance can be assigned multiple categories or a set of target labels. e.g- Genre of the movie "The Dark Knight" are Action, Crime, and Drama.

6. (1 point) What will be the output of following code snippet.

```
from sklearn.datasets import make_circles
x, y = make_circles()
#Refer polynomial_transform Function explained in the lecture.
x_poly = polynomial_transform(x, degree=3)

print (x.shape[0])
```

A. 2

B. 100

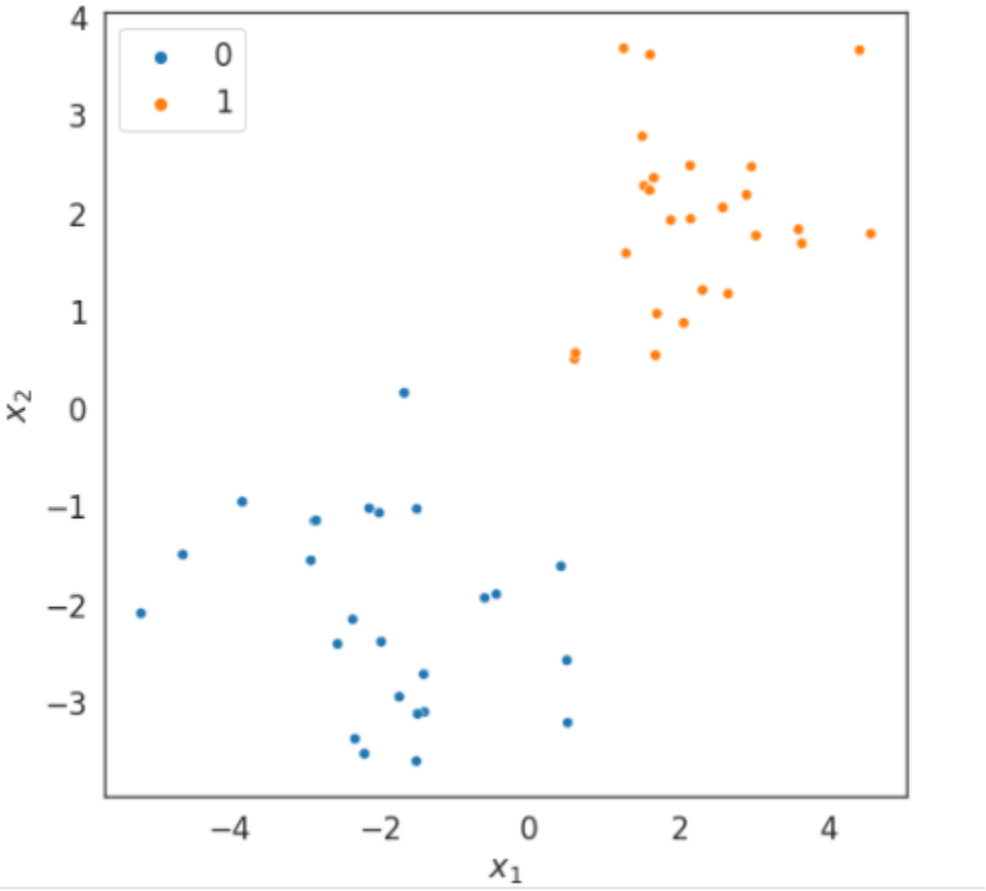
C. 6

D. 10

Answer: B

Solution: Code output is 100.

The two-dimensional chart given below is scatter plot between variable X_1 and X_2.



7. (1 point) Based on the above scatter plot, which of the following equation of the line(decision boundary) classifies orange and blue points most accurately?

   A. $X_1 + 2X_2 = 4$

   B. $X_2 - X_1 = 2$

   C. $X_1 - X_2 = 1$

   D. $X_1 + 3X_2 = 12$

   Answer: A

   Solution: We can see that Equation shown in option 1 will pass through $(0, 2)$ and $(4, 0)$ and hence the line is clearly able to classify all blue and orange points.

8. (1 point) Which of the following method adopts a dictionary-oriented approach, associating to each category label a progressive integer number.

   A. featurehasher

   B. labelencoder class

   C. labelbinarizer class

   D. dictvectorizer

   Answer: B

   Solution: Labelencoder class adopts a dictionary-oriented approach, associating to each category label a progressive integer number.

9. (1 point) Which of the following scheme is used for label encoding in multiclass setup ?

   A. One-hot encoding.

   B. Recursive Feature Elimination.

   C. Filter Method

   D. Wrapper method.

   Answer: A

   Solution: In multiclass setup we use One-hot encoding for label encoding.

10. (1 point) Multi-class classification problems have multiple categories but each instance is assigned only one of the categories.

    A. True.

B. False.

Answer: A

11. (1 point) **(Multiple Select)** Which of the following is/are examples of classification problems ?

    A. Predicting Price of airplane ticket

    B. Email classification

    C. Credit card fraud detection

    D. Malware classification

    E. Prediction of number of Umbrella sold based on the Rain happened in Area

Answer: B, C, D

Solution: A) Predicting Price of airplane ticket is regression problem.

B)In email classification we classify whether email is spammed or not.

c) In Credit card fraud detection problem our model is used to identify whether a new transaction is fraudulent or not.

D) As the name says it is a classification problem.
E) Prediction of number of Umbrella sold based on the Rain happened in Area is regression problem.