

DATAWARS PROJECTS

Learning area: Classification

Shrinivasan M

Project 1: Introduction to Classification

1 Predictions



Based on previous results, choice the correct answers.

- ☐ The predictive label of the instance 874 is 1
- ☒ The predictive label of the instance 249 is 1
- ☐ The predictive label of the instance 664 is 0
- ☒ The predictive label of the instance 874 is 0

Submitted

Correct!

Did you like this project?

2 Accuracy score



Use the following code to compute the accuracy_score:

```
from sklearn.metrics import accuracy_score

y_pred = tree.predict(X)
accuracy_score(y_pred,y)
```

Write first two digit after decimal

0.90

Submitted

Correct!

Report issue

3 Confusion Matrix



Select the correct answers using the information from the confusion matrix.

- ☒ The number of false positive instance is 61
- ☐ The model classify incorrectly 500 instance
- ☐ There are only 61 true positive and 34 true negative
- ☒ The model classify correctly 905 instances

Submitted

Correct!

Report issue

4 Create the response variable based on the columns Trump and Clinton



Separate the target and the features into two variables and create the response variable based on the columns Trump and Clinton.

Select the correct code to complete this task

There could be more than just one correct answer.

- ☐ `X=df.drop(['trump', 'clinton'],axis=1)`
- ☐ `y=df.trump`
- ☒ `y=np.where(df.trump>df.clinton,1,0)`
- ☒ `X=df[['minority','bachelor']]`

Submitted

Correct!

Report Issue

5 Decision Tree classifier



Initialize a Decision Tree classifier (name the model clf) and fit on the data with a random_state: 42 and max_depth : 3 (the maximum depth of our decision tree using the max_depth parameter).

Calculate the accuracy score of the training dataset. Select the correct answer.

- ☐ 0.901
- ☒ 0.908
- ☐ 0.801
- ☐ 0.808

Submitted

Correct!

Did you like this project?

Rate your experience



Report Issue

6

Classification

Which of the following statements are True about classification?

☒

 Classification is a supervised task.

☐

 Classification is a unsupervised task.

☒

 A single row of data is called an instance.

☒

 The target variable is the variable whose values are modeled and predicted by other variables.

Submitted

Correct!

Report issue

7

Example of classification

Which of the following statements are examples of classification?

☒

 Determine when a heater is on or off based on temperature data.

☐

 Develop a mathematical relationship between heater level (0-100%) and temperature (20-70°C).

☐

 Predict the prices of the house based on other variables

☒

 Predict when an images belong to a class such as recurrent cancer or no cancer.

Submitted

Correct!

Report issue

8

Training and testing phase

In which phase are model parameters adjusted?

☒

 Training phase

☐

 Testing phase

☐

 Data preparation phase

Submitted

Correct!

Report issue

Did you like this project?

Rate your experience

Project 2: Coding decision tree from scratch

1 Select the Gini index for the entire dataset

- ☐ 0.641
- ☐ 0.232
- ☐ 0.375
- ☒ 0.653

Submitted

Correct!

 Report Issue

2 Suppose we want to split the dataset based on color. Compute the Gini index for each subset of the dataset that results from this split.

- ☐ Gini index red fruit = 0
- ☐ Gini index blue fruit = 0.375
- ☐ Gini index blue fruit = 1
- ☒ Gini index red fruit = 0.375

Submitted

Correct!

 Report Issue

3 Which split, based on color or taste, results in a lower Gini index?

- ☒ Taste
- ☐ Color

Submitted

Correct!

Did you like this project?

Rate your experience



 Report Issue

Completed activities 4/4 100%

4 What is the Gini index of the initial split of the decision tree

- ☒ 0.5
- ☐ 0.023
- ☐ 0
- ☐ 0.025

Submitted

Correct!

 Report Issue

Project 3: K-Nearest Neighbor algorithm for classification

1 Performance evaluation



What is accuracy of training and testing data?

Round to two decimal places

- ☐ The accuracy of the test set is 0.80
- ☒ The accuracy of the test set is 0.90
- ☐ The accuracy of the train set is 0.78
- ☒ The accuracy of the train set is 0.93

Submitted

Correct!

Report issue

2 Test time versus train time



k-NN algorithm does more computation on test time rather than train time.

- ☐ No
- ☒ Yes

Submitted

Correct!

Report issue

3 KNN uses



Which of the following option is true about KNN algorithm?

- ☐ It can be used for classification
- ☐ It can be used for regression
- ☒ It can be used in both classification and regression

Submitted

Correct!

Report issue

4 Assumptions



Which of the following statement is true about KNN algorithm ?

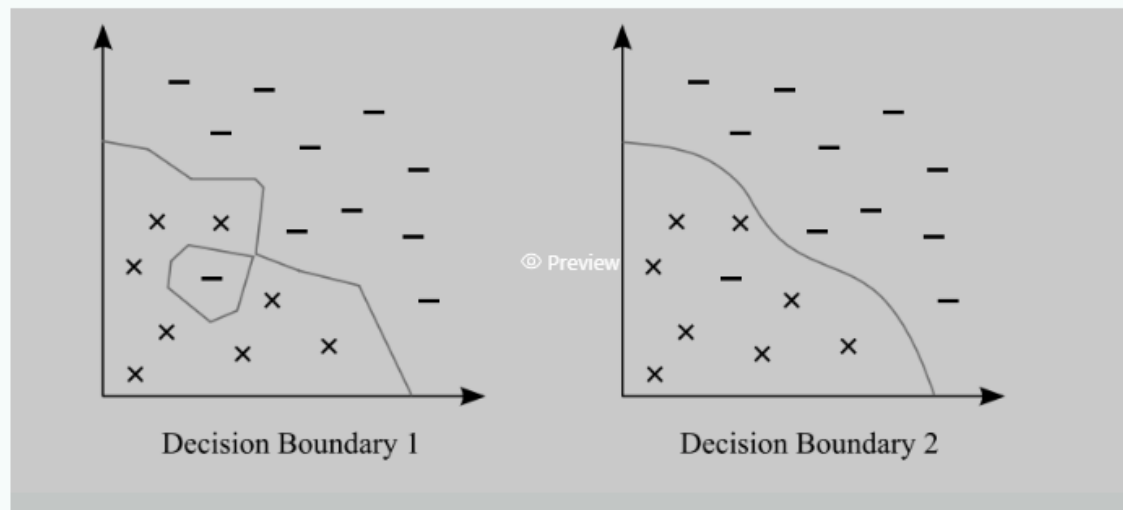
- ☒ KNN works well with a small number of input variables.
- ☐ k-NN makes assumptions about the functional form of the problem being solved.
- ☒ KNN performs much better if all of the data have the same scale.

Submitted

Correct!

5 Decision Boundaries

The Figure below illustrates decision boundaries for two nearest-neighbour classifiers. Determine which one of the boundaries belongs to the 1-nearest neighbour classifier.



- ☐ Decision boundary 2
- ☒ Decision boundary 1

Submitted

Correct!