Name: Abdul Rehman

Phone no # +923033978991

11. Can We Use The Same Data For Testing And Training, If Yes Then Why And If No, Why Not? Please Provide 1 Example That You Used To Come Up With Your Findings.

My Answer:

We can't use same data for testing and training. Sometimes the model you are training "over-learns" or memorizes the training data and performs poorly on unseen data. This is called "over fitting". The problem with training and testing on the same dataset is that we don't know that the model is over fitting because it works well on the test dataset. The purpose of testing data not visible during training is to be able to properly assess whether over fitting is occurring.

Scenarios that I captured while this ML_MARATHON

Firstly, I used RandomForestClassifier for this problem. After I trained my model I checked the accuracy of my model on training dataset:

```
In [24]: np.random.seed(42)
         model.score(x_train,y_train) #checking well model has set weights upon training dataset

Out[24]: 1.0
```

Got accuracy of 100%

But then I checked it on test dataset I got:

```
In [25]: np.random.seed(42)
         model.score(x_test,y_test)

Out[25]: 0.8465671641791045
```

This is the clear Issue of over fitting. How I came to know because there is a lot of difference in the accuracy with train and test datasets. It would be impossible to check such issues without making two different datasets