

Quiz_Answers

December 10, 2017

0.1 QUIZ-1

ANS 1: Supervised Learning

ANS 2: Unsupervised Learning

ANS 3: Classification

ANS 4: Regression

ANS 5: $k=1 \Rightarrow$ Class 1 and $k=3 \Rightarrow$ Class 2

ANS 6: Memorizes the entire training set

ANS 7:

1. See what type of cleaning or preprocessing still needs to be done
2. you might notice missing data
3. gain insight on what machine learning model might be appropriate in any
4. get a sense for how difficult the problem might be

ANS 8: to estimate how well the learned model will generalize to new data

ANS 9: To make experiment easily reproducible by always using the same partitioning of the data

ANS 10:

```
X_train (7500,50)
y_train(7500,)
X_test(2500,50)
y_test(2500,)
```

0.2 QUIZ-2

ANS 1: You are overfitting, the next model trained should have a higher value for alpha

ANS 2: Decrease C and gamma

ANS 3: Classify a set of fruits as apples, oranges, bananas, or lemons

ANS 4: 10

ANS 5: Lasso Regression

ANS 6: 0.1, 1, 10

ANS 7: Figure A: Ridge Regression, Figure B: Lasso Regression

ANS 8: 3

ANS 9: 10

ANS 10: 8

ANS 11:

1. Helps prevent knowledge about the test set from leaking into the model
2. Increases generalization ability and computational complexity
3. Fits multiple models on different splits of the data

0.3 QUIZ-3

ANS 1: 0.99

ANS 2: 0.906

ANS 3: 0.923

ANS 4: 0.960

ANS 5: 0.6

ANS 6:

1. Model 1:Roc 1

2. Model 2:Roc 3

3. Model 3:Roc 2

ANS 7: Not enough information is given

ANS 8: 0.744

ANS 9:

1. The best possible score is 1.0
2. The worst possible score is 0.0

ANS 10: Precision

ANS 11: Recall

ANS 12: The model is probably misclassifying the frequent labels more than the infrequent labels.

ANS 13: 0.52

ANS 14: 0.15

0.4 QUIZ-4

ANS 1: Separate the data into distinct groups by similarity

ANS 2:

1. Trees often require less preprocessing of data
2. Trees are easy to interpret and visualize

ANS 3: To improve generalization by reducing correlation among the trees and making the model more robust to bias.

ANS 4: Neural Network, SVM, KNN

ANS 5:

1. For a model that won't overfit a training set, Naive Bayes would be a better choice than
2. For predicting future sales of a clothing line, Linear regression would be a better choice than a regressor.

ANS 6: Neural Network -> KNN(k=1) -> Decision Tree

ANS 7: 0.06745

ANS 8:

1. collection_status - Flag for payments in collections
2. compliance_detail - More information on why each ticket was marked compliant or non-compliant

ANS 9:

1. If time is a factor, remove any data related to the event of interest that doesn't take place at the same time
2. Remove variables that a model in production wouldn't have access to
3. Sanity check the model with an unseen validation set

ANS 10: EXOR (output would be 0110)

In []: