ML2 Mock Test https://bit.ly/2NzgWQN

1) Which of the following metrics does NOT penalize the (regression) model for having extra features in it?

- AIC
- BIC
- MSE
- Adjusted R-squared

2) The kernel trick in SVM involves:

- Pairwise dot-products of feature vectors expressed as a simple scalar function over the original attribute vectors
- Transforming data points to a higher dimensional space where the separator is likely to be linear
- Optimizing using a more complex function of the attributes and the model as the margin
- None of the above

3) Which of the following is true about ridge and lasso regression methods?

- Ridge regression performs subset selection of features
- Lasso regression performs subset selection of features
- Both perform subset selection of features
- None of the above

4) The maximum margin classifier is also:

- The classifier that is immune to data points that are not support vectors
- The model that minimizes the 2-norm of the coefficient vector
- The classifier that generalizes best given the current training data
- All of the above

5) SVM is a sparse model because:

- Most of the weights in the SVM model end up becoming zero
- Most of the training data points effectively become irrelevant
- The feature vector corresponding to the kernel is sparse
- None of the above

6) SVM generalizes well largely because

- It is a linear model
- Allows a general class of kernels to be used
- It maximizes the (separation) margin on the dataset
- None of the above

7) If a hard margin SVM model shows a low training and validation accuracy then the corrective action, assuming the same data will involve:

- Better cross-validation
- Choosing a more powerful kernel
- Selecting better features
- None of the above

8) The decision tree for a binary labelled dataset happens to have one label (say 1) strongly correlated with one value of an attribute A and another label (0) strongly correlated with one value of another attribute B. The depth and the number of leaves of the decision tree for the dataset is likely to be:

- 2, 2
- 2, 3
- 1, 3
- 3, 2

9) Cross Validation results indicate a good choice of the model when it shows

- High average accuracy with high variance
- Low average accuracy with low variance
- High average accuracy with low variance
- Low average accuracy with high variance

10) The primary issue with models having high variance is:

- Model produced is too sensitive to the input data – it will also learn what shouldn't be learnt from the data given
- Model becomes too complex to deal with
- Takes a lot of time and computational effort to train such models
- All of the above

11) Variance is not suitable for use as a homogeneity measure for classification problems in a decision tree, because:

- It will not give good results
- Variance can be computed only for real-valued labels
- Class labels are not numeric
- Class labels are usually unordered and there is no 'distance' defined between classes.

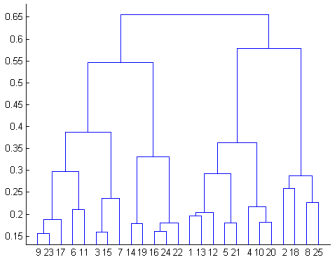
12) Which of the following is not directly 'inherited' by a Random Forest from a decision tree?

- Robustness against missing values and different attribute types
- Not affected much by outliers and no need for explicit normalization
- The relative importance of features and interpretability in terms of rules
- None of the above

13) More the merrier for trees in a random forest holds as long as

- There are lots of features to choose from for the split
- Most of the trees in the forest have a low validation error
- There is low pairwise correlation among trees in the forest
- The dataset is large

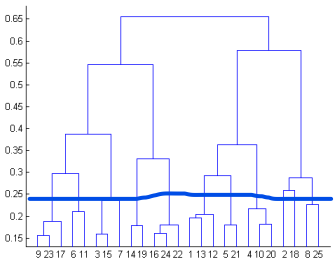
14) Refer to the dendrogram image below and answer the 2 questions that follow:



Which of the following 2 points will get clustered in the first step of hierarchical clustering?

- 6,11
- 24, 22
- 9, 23
- 5, 21

15) Find the number of clusters formed if the dendrogram is cut at 0.25. (Assume agglomerative clustering method):



- 6
- 11
- 13
- 15

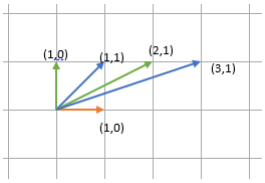
16) Consider the following statements.

a) The principal components are linear combinations of the original variables
b) All principal components are orthogonal to each other
c) PCA won't work well in a dataset that is highly correlated

Which of the above statement(s) are NOT TRUE?

- Only A
- Only C
- A and B
- A,B & C

17) In the following diagram, the orange vector is given by (1,0), the blue vector by (1,1) and the green vector by (0,1). The transformation matrix T = Failed to load image. Click here to try again. is applied which sends the three vectors to the orange vector (1,0), the blue vector (3,1) and the green vector (2,1), respectively.



Which of the three original vectors is an eigenvector of the linear transformation T?

- (1,0)
- (1,1)
- (0,1)
- None of these