# Solved Machine Learning

1. The value of correlation coefficient will always be:
   A) between 0 and 1
   B) greater than -1
   C) between -1 and 1
   D) between 0 and -1
   Ans- C) between -1 and 1

2. Which of the following cannot be used for dimensionality reduction?
   A) Lasso Regularisation
   B) PCA
   C) Recursive feature elimination
   D) Ridge Regularisation
   Ans- D) Ridge Regularisation

3. Which of the following is not a kernel in Support Vector Machines?
   A) linear
   B) Radial Basis Function
   C) hyperplane
   D) polynomial
   Ans- A) linear

4. Amongst the following, which one is least suitable for a dataset having non-linear decision boundaries?
   A) Logistic Regression
   B) Naïve Bayes Classifier
   C) Decision Tree Classifier
   D) Support Vector Classifier
   Ans- B) Naïve Bayes Classifier

5. In a Linear Regression problem, ‘X’ is independent variable and ‘Y’ is dependent variable, where ‘X’ represents weight in pounds. If you convert the unit of ‘X’ to kilograms, then new coefficient of ‘X’ will be? (1 kilogram = 2.205 pounds)

   A) 2.205 × old coefficient of ‘X’
   B) same as old coefficient of ‘X’
   C) old coefficient of ‘X’ ÷ 2.205
   D) Cannot be determined

   Ans- A) 2.205 × old coefficient of ‘X’

6. As we increase the number of estimators in ADABOOST Classifier, what happens to the accuracy of the model?

   A) remains same
   B) increases
   C) decreases
   D) none of the above

   Ans- C) decreases

7. Which of the following is not an advantage of using random forest instead of decision trees?

   A) Random Forests reduce overfitting
   B) Random Forests explains more variance in data then decision trees
   C) Random Forests are easy to interpret
   D) Random Forests provide a reliable feature importance estimate

   Ans- B) Random Forests explains more variance in data then decision trees

8. Which of the following are correct about Principal Components?

A) Principal Components are calculated using supervised learning techniques

B) Principal Components are calculated using unsupervised learning techniques

C) Principal Components are linear combinations of Linear Variables.

D) All of the above

Ans- B) Principal Components are calculated using unsupervised learning techniques

C) Principal Components are linear combinations of Linear Variables.

9. Which of the following are applications of clustering?

A) Identifying developed, developing and under-developed countries on the basis of factors like GDP, poverty index, employment rate, population and living index

B) Identifying loan defaulters in a bank on the basis of previous years' data of loan accounts.

C) Identifying spam or ham emails

D) Identifying different segments of disease based on BMI, blood pressure, cholesterol, blood sugar levels.

Ans- A) Identifying developed, developing and under-developed countries on the basis of factors like GDP, poverty index, employment rate, population and living index

C) Identifying spam or ham emails

10. Which of the following is(are) hyper parameters of a decision tree?

A) max_depth

B) max_features

C) n_estimators

D) min_samples_leaf

Ans- A) max_depth

B) max_features

D) min_samples_leaf

11. What are outliers? Explain the Inter Quartile Range (IQR) method for outlier detection.

Ans.

Outliers are those data points that are significantly different from the rest of the dataset. They are often abnormal observations that skew the data distribution, and arise due to inconsistent data entry, or erroneous observations.

IQR (Interquartile Range) is used to measure variability by dividing a data set into quartiles. The data is sorted in ascending order and split into 4 equal parts. Q1, Q2, Q3 called first, second and third quartiles are the values which separate the 4 equal parts.

- Q1 represents the 25th percentile of the data.
- Q2 represents the 50th percentile of the data.
- Q3 represents the 75th percentile of the data.

If a dataset has *2n / 2n+1* data points, then
Q1 = median of the dataset.
Q2 = median of n smallest data points.
Q3 = median of n highest data points.
IQR is the range between the first and the third quartiles namely Q1 and Q3: *IQR = Q3 – Q1*. The data points which fall below *Q1 – 1.5 IQR* or above *Q3 + 1.5 IQR* are outliers.

---

12. What is the primary difference between bagging and boosting algorithms?

Ans.

The primary differences between bagging and boosting algorithms are:

- Bagging is the simplest way of combining predictions that belong to the same type while Boosting is a way of combining predictions that belong to the different types.
- Bagging aims to decrease variance, not bias while Boosting aims to decrease bias, not variance.
- In Baggiing each model receives equal weight whereas in Boosting models are weighted according to their performance.
- In Bagging each model is built independently whereas in Boosting new models are influenced by performance of previously built models.
- In Bagging different training data subsets are randomly drawn with replacement from the entire training dataset. In Boosting every new

subsets contains the elements that were misclassified by previous models.

- Bagging tries to solve over-fitting problem while Boosting tries to reduce bias.
- If the classifier is unstable (high variance), then we should apply Bagging. If the classifier is stable and simple (high bias) then we should apply Boosting.
- Bagging is extended to Random forest model while Boosting is extended to Gradient boosting.

---

13. What is adjusted R2 in linear regression. How is it calculated?

Ans

Adjusted $R^2$ is a corrected goodness-of-fit (model accuracy) measure for linear models. It identifies the percentage of variance in the target field that is explained by the input or inputs. Adjusted R squared is calculated by dividing the residual mean square error by the total mean square error (which is the sample variance of the target field). The result is then subtracted from 1. Adjusted $R^2$ is always less than or equal to $R^2$. A value of 1 indicates a model that perfectly predicts values in the target field. A value that is less than or equal to 0 indicates a model that has no predictive value. In the real world, adjusted $R^2$ lies between these values.

---

14. What is the difference between standardisation and normalisation?

Ans

The difference between standardization and normalization-

| Standardization | Normalization |
|---|---|
| Minimum and maximum value of features are used for scaling | Mean and standard deviation is used for scaling. |
| It is used when features are of different scales | It is used when we want to ensure zero mean and unit standard deviation |
| Scales values between [0, 1] or [-1, 1]. | It is not bounded to a certain range. |
| It is really affected by outliers | It is much less affected by outliers |
| Scikit-Learn provides a transformer called MinMaxScaler for Normalization. | Scikit-Learn provides a transformer called StandardScaler for standardization. |
| This transformation squishes the n-dimensional data into an n-dimensional unit hypercube | It translates the data to the mean vector of original data to the origin and squishes or expands |
| It is useful when we don't know about the distribution | It is useful when the feature distribution is Normal or Gaussian. |
| It is often called as Scaling Normalization | It is often called as Z-score Normalization |

---

15. What is cross-validation? Describe one advantage and one disadvantage of using cross-validation.

Ans

Cross-validation is a statistical method used to estimate the performance (or accuracy) of machine learning models. It is used to protect against overfitting in a predictive model, particularly in a case where the amount of data may be limited. In cross-validation, you make a fixed number of folds (or partitions) of the data, run the analysis on each fold, and then average the overall error estimate.

Advantage – Use all the data- When we have very little data, splitting it into training and test set might leave us with a very small test set. Say we have

only 100 examples, if we do a simple 80–20 split, we'll get 20 examples in our test set. It is not enough. We can get almost any performance on this set only due to chance. The problem is even worse when we have a multi-class problem. If we have 10 classes and only 20 examples, It leaves us with only 2 examples for each class on average. Testing anything on only 2 examples can't lead to any real conclusion. If we use cross-validation in this case, we build K different models, so we are able to make predictions on all of our data. For each instance, we make a prediction by a model that didn't see this example, and so we are getting 100 examples in our test set. For the multi-class problem, we get 10 examples for each class on average, and it's much better than just 2. After we evaluated our learning algorithm (see #2 below) we are now can train our model on all our data because if our 5 models had similar performance using different train sets, we assume that by training it on all the data will get similar performance. By doing cross-validation, we're able to use all our 100 examples both for training and for testing while evaluating our learning algorithm on examples it has never seen before.

Disadvantage- Increases Training Time: Cross Validation drastically increases the training time. Earlier you had to train your model only on one training set, but with Cross Validation you have to train your model on multiple training sets. For example, if you go with 5 Fold Cross Validation, you need to do 5 rounds of training each on different 4/5 of available data. And this is for only one choice of hyperparameters. If you have multiple choice of parameters, then the training period will shoot too high