MACHINE LEARNING

ASSIGNMENT - 5

Q1 to Q15 are subjective answer type questions, Answer them briefly.

Q1. R-squared or Residual Sum of Squares (RSS) which one of these two is a better measure of goodness of fit model in regression and why?

ANS: R2: it represents the proportion of the variance in our data which is explained by our model; the closer to one, the better the fit. The residual sum of squares (RSS) is the sum of the squared distances between actual versus predicted values:
RSS=∑i=1 to n [(yi – y'i)^2]
Where yi is a given datapoint and y'i is the fitted value for yi. The actual number we get depends largely on the scale of our response variable. Taken alone, the RSS isn't so informative.
Therefore, R2 is a better measure.

Q2. **What are TSS (Total Sum of Squares), ESS (Explained Sum of Squares) and RSS (Residual Sum of Squares) in regression. Also mention the equation relating these three metrics with each other.**

ANS: The residual sum of squares (RSS) is the sum of the squared distances between actual versus predicted values:
RSS=∑i=1 to n [(yi – y'i)^2]
ESS: The explained sum of squares (ESS) is the sum of the squares of the deviations of the predicted values from the mean value of a response variable, in a standard regression model.
ESS=∑i=1 to n [(y'i – ymean)^2]
TSS: Total sum of squares (TSS) = explained sum of squares (ESS)+ residual sum of squares (RSS).
TSS = ∑i=1 to n [(y'i – ymean)^2] + ∑i=1 to n [(yi – y'i)^2]

Q3. What is the need of regularization in machine learning?

ANS: Regularisation is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting.
Regularization significantly reduces the variance of the model, without substantial increase in its bias

Q4. What is Gini–impurity index?

ANS: Gini Index, also known as Gini impurity, **calculates the amount of probability of a specific feature that is classified incorrectly when selected randomly.** If all the elements are linked with a single class then it can be called pure.

Let's perceive the criterion of the Gini Index, like the properties of entropy, ***the Gini index varies between values 0 and 1,*** *where 0 expresses the purity of classification, i.e. All the elements belong to a specified class or only one class exists there. And 1 indicates the random distribution of elements across various classes. The value of 0.5 of the Gini Index shows an equal distribution of elements over some classes.*

While designing the decision tree, the features possessing the least value of the Gini Index would get preferred. You can learn another tree-based algorithm(Random Forest).

The Gini Index is determined by deducting the sum of squared of probabilities of each class from one, mathematically, Gini Index can be expressed as:

$$\text{Gini Index} = 1 - \sum_{i=1}^{n} (P_i)^2$$

Gini Index Formula

Q5. Are unregularized decision-trees prone to overfitting? If yes, why?

ANS: Yes, unregularized decision trees are prone to overfitting. Decision trees are prone to overfitting, especially when a tree is particularly deep. This is due to the amount of specificity we look at leading to smaller sample of events that meet the previous assumptions. This small sample could lead to unsound conclusions.

6. What is an ensemble technique in machine learning?

ANS: In this ensemble technique, **machine learning professionals use a number of models for making predictions about each data point**. The predictions made by different models are taken as separate votes. Subsequently, the prediction made by most models is treated as the ultimate prediction.

Ensemble techniques combine the decisions from multiple models to improve the overall performance. Bagging and Boosting are two of the most used techniques in machine learning

Q7. What is the difference between Bagging and Boosting techniques?

ANS : Bagging is a homogeneous weak learners' model that learns from each other independently in parallel and combines them for determining the model average.

• Boosting is also a homogeneous weak learners' model but works differently from Bagging. In this model, learners learn sequentially and adaptively to improve model predictions of a learning algorithm

Q8. What is out-of-bag error in random forests?

ANS : The out-of-bag (OOB) error is **the average error for each calculated using predictions from the trees that do not contain in their respective bootstrap sample**

Q9. What is K-fold cross-validation?

ANS: In k-fold cross-validation, the original sample is randomly partitioned into k equal sized subsamples. Of the k subsamples, a single subsample is retained as the validation data for testing the model, and the remaining k − 1 subsamples are used as training data. The cross-validation process is then repeated k times, with each of the k subsamples used exactly once as the validation data. The k results can then be averaged to produce a single estimation.

Q10. What is hyper parameter tuning in machine learning and why it is done?

ANS: In machine learning, hyperparameter optimization or tuning is **the problem of choosing a set of optimal hyperparameters for a learning algorithm**. A hyperparameter is a parameter whose value is used to control the learning process. By contrast, the values of other parameters (typically node weights) are learned.

Q11. What issues can occur if we have a large learning rate in Gradient Descent?

ANS:A learning rate that is too large **can cause the model to converge too quickly to a suboptimal solution**, whereas a learning rate that is too small can cause the process to get stuck. The challenge of training deep learning neural networks involves carefully selecting the learning rate

Q12. Can we use Logistic Regression for classification of Non-Linear Data? If not, why?

ANS: It can only be used to predict discrete functions. Hence, the dependent variable of Logistic Regression is bound to the discrete number set. It is very fast at classifying unknown records. **Non-linear problems can't be solved with logistic regression because it has a linear decision surface**.

Q13. Differentiate between Adaboost and Gradient Boosting.

ANS: Gradient Boosting is a generic algorithm to find approximate solutions to the additive modelling problem, while AdaBoost be a special case with a particular loss function. Hence, Gradient Boosting is much more flexible.

On the other hand, AdaBoost can be interpreted from a much more intuitive perspective and can be implemented without the reference to gradients by reweighting the training samples based on classifications from previous learners

Q14. What is bias-variance trade off in machine learning?

ANS: If our model is too simple and has very few parameters then it may have high bias and low variance. On the other hand if our model has large number of parameters then it’s going to have high variance and low bias. So we need to find the right/good balance without overfitting and underfitting the data.

This tradeoff in complexity is why there is a tradeoff between bias and variance. An algorithm can’t be more complex and less complex at the same time.

Q15. Give short description each of Linear, RBF, Polynomial kernels used in SVM.?

ANS: • Linear Kernel is used when the data is Linearly separable, that is, it can be separated using a single Line. It is one of the most common kernels to be used. It is mostly used when there are many Features in a particular Data Set.

• Gaussian RBF(Radial Basis Function) is another popular Kernel method used in SVM models for more. RBF kernel is a function whose value depends on the distance from the origin or from some point. Gaussian Kernel is of the following format

• In the polynomial kernel, we simply calculate the dot product by increasing the power of the kernel.