MACHINE LEARNING WORKSHEET-5

ANSWERS

1. 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:-**R-Squared** is a statistical measure which represents the percentage of the variance in the dependent variable that the independent variables explain collectively. The strength of the relationship between the model and the dependent variable on a convenient 0 – 100% scale.

The **Residual sum of squares (RSS)** ,also known as sum of squared residuals(SSR)is a statistical technique used to measure the amount of variance in a data set that is not explained by a regression model. Instead, it estimates the variance in the residuals, or error term.

The smaller the RSS, the better your model fits your data; the greater the RSS ,the poorer your model fits your data. A value of zero in RSS means your model is a perfect fit.

Moreover, **R-squared measures the goodness-of-fit model for linear regression models much better than RSS.** R-Squared explains to what extent the variance of one variable explains the variance of second variable. So , if the R-Squared of a model is 0.50,then approximately half of the observed variation can be explained by model's inputs.

2. 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:- **Residual sum of square (RSS)** is given by the summation of squares of error values i.e ground value , predicted value .

Explained sum of squares(ESS) is given by the summation of squares of the deviation of the predicted value from the mean of the variable .

Total sum of squared(TSS) is given by the summation of deviation of ground truth from the mean of the variable .

The relation between these above three could be linearly expressed as :-

TSS=RSS+ESS

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

ANS**:-** **Regularization** refers to techniques that are used to calibrate machine learning models in order to minimize the adjusted loss function and prevent overfitting or underfitting.

- Using Regularization, we can fit our machine learning model appropriately on a given test set and hence reduce the errors in it. This means that this technique discourages learning a more complex or flexible model, so as to avoid the risk of overfitting.

There are two main types of regularization techniques:-

I. Ridge Regularization
II. Lasso Regularization.

4. 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.

This measures is calculated where the model contains Decision Tree Algorithms or Random Forest algorithms.

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

Ans:-Yes ,Decision tree are prone to overfitting, If it is allowed to train to its full strength, the model will overfit the training data. But unlike other algorithms, decision tree does not use regularization to fight against overfitting.

There are various techniques to prevent the decision tree model from overfitting. These techniques are :

- **Pruning**— Pruning refers to a technique to remove the parts of the decision tree to prevent growing to its full depth.
  --**pre pruning** : pre pruning will stop growing the tree earlier, before it perfectly classifies the training set.
  --**post pruning**: that allows the tree to perfectly classify the training set, and then post prune the tree.

6. What is an ensemble technique in machine learning?

Ans: - **Ensemble methods** are techniques that aim at improving the accuracy of results in models by combining multiple models instead of using a single model. The combined models increase the accuracy of the results significantly. This has boosted the popularity of ensemble methods in machine learning.

- Ensemble methods are ideal for regression and classification, where they reduce bias and variance to boost the accuracy of models.

- Random Forest, XG Boosts, Gradient Boosting are some examples of ensemble learning techniques.

- There are 2 types of ensemble techniques: -
  I. Bagging
  II. Boosting

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

Ans:- **Bagging--B**ootstrap **A**ggregating, also known as bagging, is a machine learning ensemble meta-algorithm designed to improve the stability and accuracy of machine learning algorithms used in statistical classification and regression.

It decreases the variance and helps to avoid overfitting. It is usually applied to decision tree methods. Bagging is a special case of the model averaging approach.

The Random Forest model uses Bagging, where decision tree models with higher variance are present.

Boosting is an ensemble modeling technique that attempts to build a strong classifier from the number of weak classifiers. It is done by building a model by using weak models in series.

- Firstly, a model is built from the training data.
- Then the second model is built which tries to correct the errors present in the first model.
- This procedure is continued and models are added until either the complete training data set is predicted correctly or the maximum number of models is added.

The difference between bagging and boosting techniques are as follows :-

S.No	BAGGING	BOOSTING
1.	The simplest way of combining predictions that belong to the same type.	A way of combining predictions that belong to the different types.
2.	Aim to decrease variance, not bias.	Aim to decrease bias, not variance.
3.	Each model receives equal weight.	Models are weighted according to their performance.
4.	Bagging tries to solve the over-fitting problem.	Boosting tries to reduce bias.

5.	In these base classifiers are trained parallelly.	In this base classifier are trained sequentially.
6.	If the classifier is unstable (high variance), then apply bagging.	If the classifier is stable and simple (high bias) the apply boosting.
Example:	The Random Forest model uses Bagging.	**Example:** The AdaBoost uses Boosting techniques

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

Ans:- **OOB (out-of-bag) errors** are an estimate of the performance of a random forest classifier or regressor on unseen data. In scikit-learn, the OOB error can be obtained using the oob_score_ attribute of the random forest classifier or regressor.

The OOB error is computed using the samples that were not included in the training of the individual trees. OOB error is not always a reliable estimate of the generalization error of the model, but it can provide a useful indication of how well the model is performing.

9. What is K-fold cross-validation?

Ans:- **Cross-validation** is a resampling procedure used to evaluate machine learning models on a limited data sample.

- This procedure has **a single parameter called k** that refers to the number of groups that a given data sample is to be split into and is often **called k-fold cross-validation**.
- When a specific value for k is chosen, such as k=10 becoming 10-fold cross-validation.
  Cross-validation is primarily used in applied machine learning to estimate the skill of a machine learning model on unseen data. It use a limited sample in order to estimate how the model is expected to perform when used to make predictions on data which is not used during the training of the model.
- It is a popular method because it is simple to understand and because it generally results in a less biased or less optimistic estimate of the model skill than other methods, such as a simple train/test split.

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

Ans:- A hyperparameter is a parameter of the model whose value influences the learning process and whose value cannot be estimated from the training data. Hyperparameters are configured externally before starting the model learning/training process.

- **Hyperparameter tuning** is the process of finding the optimal hyperparameters for any given machine learning algorithm.

Hyperparameter-tuning is important **to find the possible best sets of hyperparameters to build the model from a specific dataset**. There are two main types of hyperparameter tuning:

1) **Manual hyperparameter tuning**
2) **Automated hyperparameter tuning**

The two best strategies for Hyperparameter tuning are:

- Grid Search CV
- Randomized Search CV

11. 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.

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

Ans:-Logistic regression has been used as a linear classifier, i.e., when the classes can be separated by linear boundaries. Logistic regression is known and used as a linear classifier.

13. Differentiate between Adaboost and Gradient Boosting.?

Ans:- **Gradient Boosting** is the boosting algorithm that works on the principle of the stagewise addition method, where multiple weak learning algorithms are trained and a strong learner algorithm is used as a final model from the addition of multiple weak learning algorithms trained on the same dataset.

On the other hand, **AdaBoost** is a boosting algorithm, which also works on the principle of the stagewise addition method where multiple weak learners are used for getting strong learners.

→ Unlike Gradient Boosting in XGBoost, the alpha parameter is related to the errors of the weak learner, here the value of the alpha parameter will be indirectly proportional to the error of the weak learner.

The difference between Adaboost and Gradient boosting techniques are as follows:-

S.No	ADABOOST	GRADIENT BOOSTING
1.	An additive model where shortcomings of previous models are identified by high-weight data points.	An additive model where shortcomings of previous models are identified by the gradient.
2.	The trees are usually grown as decision stumps.	The trees are grown to a greater depth usually ranging from 8 to 32 terminal nodes.
3.	It gives weights to both classifiers and observations thus capturing maximum variance within data.	It builds trees on previous classifier's residuals thus capturing variance in data.

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

Ans:-In statistics and machine learning, the **bias–variance tradeoff** is the property of a model that the variance of the parameter estimated across samples can be reduced by increasing the bias in the estimated parameters.

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

Ans:- **Linear SVM is** used for linearly separable data, which means if a dataset can be classified into two classes by using a single straight line, then such data is termed as linearly separable data, and classifier is used called as Linear SVM classifier.

Ex:-
```
from sklearn.svm import SVC
classifier = SVC(kernel ='linear')
classifier.fit(x_train, y_train) # training set in x, y axis
```

RBF(Radial Basis Function) kernel SVM is a very powerful kernel used in SVM. Unlike linear or polynomial kernels, RBF is more complex and efficient at the same time that it can combine multiple polynomial kernels multiple times of different degrees to project the non-linearly separable data into higher dimensional space so that it can be separable using a hyperplane.

Ex:-
```
from sklearn.svm import SVC
classifier = SVC(kernel ='rbf', random_state = 0)
classifier.fit(x_train, y_train)
```

Polynomial kernels in SVM represents the similarity of vectors in the training set of data in a feature space over polynomials of the original variables used in the kernel.

Ex:-
```
from sklearn.svm import SVC
classifier = SVC(kernel ='poly', degree = 4)
classifier.fit(x_train, y_train) # training set in x, y axis
```