| | |
|---|---|
| **Name of the Assignment** | : Machine Learning (Worksheet-5) |
| **Submitted by** | : Shalini Joshi |
| **Designation** | : Data Science Intern |
| **Date of Submission** | : 19th Feb,2023 |

**Subjective answer type questions:**

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^2** means: it represents the proportion of the variance in your data which is explained by your model; the closer to one, the better the fit.
The **residual sum of squares (RSS)** is the sum of the squared distances between your actual versus your predicted values:

**RSS=∑i=1n(yi−y^i)2**
Where yi's a given datapoint and y^i your fitted value for yi
The actual number you get depends largely on the **scale** of your response variable. Taken alone, the RSS isn't so informative.

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: Explained sum of squares (ESS)**: Also known as the explained variation, the ESS is the portion of total variation that measures how well the regression equation explains the relationship between X and Y.

You compute the ESS with the formula:

$$ESS = \sum_{i=1}^{n}\left(\hat{Y}_i - \bar{Y}\right)^2$$

**Residual sum of squares (RSS):** This expression is also known as unexplained variation and is the portion of total variation that measures discrepancies (errors) between the actual values of Y and those estimated by the regression equation.

You compute the RSS with the formula:

$$RSS = \sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2$$

**Total sum of squares (TSS):** The Total SS (TSS or SST) tells you how much variation there is in the dependent variable.
Total SS = Σ(Yi – mean of Y)2.

The sum of RSS and ESS equals TSS. The equation relating these three metrics with each other:

**TSS =RSS+ESS**

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

**Ans:** One can easily overfit or underfit a machine learning model while it is being trained. To prevent this, we use regularization in machine learning to accurately fit a model onto our test set.

**Regularization** is a technique used to reduce the errors by fitting the function appropriately on the given training set and avoid overfitting.

In order to obtain the best model, regularization techniques help us lower the likelihood of overfitting.

4. What is Gini–impurity index?

**Ans:** Gini Impurity is a measurement used to build Decision Trees to determine how the features of a dataset should split nodes to form the tree. More precisely, the Gini Impurity of a dataset is a number between 0-0.5, which indicates the likelihood of new, random data being misclassified if it were given a random class label according to the class distribution in the dataset.

The Gini Index or Gini Impurity is calculated by subtracting the sum of the squared probabilities of each class from one. It favours mostly the larger partitions and are very simple to implement. In simple terms, it calculates the probability of a certain randomly selected feature that was classified incorrectly.

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

**Ans:** In the case of decision tree's, they can learn a training set to a point of high granularity that makes them easily overfit. Allowing a decision tree to split to a granular degree, is the behavior of this model that makes it prone to learning every point extremely well — to the point of perfect classification — ie: overfitting.

6. What is an ensemble technique in machine learning?

**Ans: Ensemble technique** is a machine learning technique that combines several base models in order to produce one optimal predictive model. This type of machine learning algorithm helps in improving the overall performance of the model.

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

**Ans:** Differences between Bagging and Boosting techniques:

| S.No. | Bagging | Boosting |
|---|---|---|
| 1 | The simplest way of combining predictions that belong to the same type. | A way of combining predictions that |
| 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 | Each model is built independently. | New models are influenced by the performance of previously built models. |
| 5 | Bagging tries to solve the over-fitting problem. | Boosting tries to reduce bias. |
| 6 | Example: The Random forest model uses Bagging. | Example: The AdaBoost uses Boosting techniques |

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

**Ans:** The **out-of-bag error** is the average error for each predicted outcome calculated using predictions from the trees that do not contain that data point in their respective bootstrap sample. This way, the Random Forest model is constantly being validated while being trained

9. What is K-fold cross-validation?

**Ans: k-Fold Cross-Validation** is a resampling procedure used to evaluate machine learning models on a limited data sample. The procedure has a single parameter called k that refers to the number of groups that a given data sample is to be split into.

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

**Ans:** In machine learning, tuning or hyper parameter optimization is the difficulty of picking a collection of optimal parameters for a model learning algorithm. A hyper Parameter is also called a model predictor, since its value is used as a starting point for the model learning algorithm.

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

**Ans:** When the learning rate is too large, gradient descent can inadvertently increase rather than decrease the training error. When the learning rate is too small, training is not only slower, but may become permanently stuck with a high training error.

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

**Ans:** Logistic regression is known and used as a linear classifier. It is used to come up with a hyper*plane* in feature space to separate observations that belong to a class from all the other observations that do *not* belong to that class. The decision boundary is thus *linear*.

While logistic regression makes core assumptions about the observations such as IID (each observation is independent of the others and they all have an identical probability distribution), the use of a linear decision boundary is *not* one of them. The linear decision boundary is used for reasons of simplicity following the Zen mantra – when in doubt simplify. In those cases where we suspect the decision boundary to be nonlinear, it may make sense to formulate logistic regression with a nonlinear model and evaluate how much better we can do.

13. Differentiate between Adaboost and Gradient Boosting.

**Ans:** AdaBoost is the first designed boosting algorithm with a particular loss function. On the other hand, Gradient Boosting is a generic algorithm that assists in searching the approximate solutions to the additive modelling problem. This makes Gradient Boosting more flexible than AdaBoost.

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

**Ans:** The **bias-variance tradeoff** refers to the tradeoff that takes place when we choose to lower bias which typically increases variance, or lower variance which typically increases bias.

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

**Ans:** SVM algorithms use a set of mathematical functions that are defined as the kernel. The function of kernel is to take data as input and transform it into the required form. Different SVM algorithms use different types of kernel functions. These functions can be different types.

- **Polynomial Kernel**: It is popular in image processing.
  Equation is:

$$k(\mathbf{x_i}, \mathbf{x_j}) = (\mathbf{x_i} \cdot \mathbf{x_j} + 1)^d$$

- **Gaussian radial basis function (RBF)**

  It is a general-purpose kernel; used when there is no prior knowledge about the data.
  Equation is:

$$k(\mathbf{x_i}, \mathbf{x_j}) = \exp(-\gamma \|\mathbf{x_i} - \mathbf{x_j}\|^2)$$

Gaussian radial basis function (RBF)

for:

$$\gamma > 0$$

- **Linear splines kernel in one-dimension**

  It is useful when dealing with large sparse data vectors. It is often used in text categorization.

The splines kernel also performs well in regression problems. Equation is:

$$k(x,y) = 1 + xy + xy\ min(x,y) - \frac{x+y}{2}\ min(x,y)^2 + \frac{1}{3}\min(x,y)^3$$

**********************  **The End**  *****************************