Week 12 Deliverables

- Name Manoj Nagaraja
- Email manojn.7270@gmail.com
- Country United Kingdom
- College University of Liverpool
- Specialization Data Science

- Github Repo Link https://github.com/ManojN7270/Final-Project-week_7-to-week_13.git

Base Model: Logistic Regression

- Logistic Regression is a linear model used for classification tasks.
- It models the probability of the input belonging to a certain class using the logistic function.
- Logistic Regression is computationally efficient and provides interpretable results.
- It can handle both binary and multi-class classification problems.
- However, it assumes a linear relationship between the features and the log-odds of the target variable.

Linear Model: Support Vector Machines (SVM)

- SVM is a versatile linear model that can be used for both classification and regression tasks.
- It aims to find a hyperplane that separates the data into different classes with the largest margin.
- SVM can handle high-dimensional data and is effective when the data is linearly separable.
- By using different kernel functions (e.g., polynomial, radial basis function), SVM can handle non-linear separation as well.

Linear models are widely used for classification tasks due to their simplicity and interpretability. One popular linear model is Logistic Regression, which models the relationship between the features and the probability of belonging to a specific class. It is a powerful and efficient algorithm for binary classification problems.

Ensemble Model: Random Forest

- Random Forest is an ensemble learning method that combines multiple decision trees.
- It creates a forest of decision trees by training each tree on a random subset of the data and features.
- Random Forest provides robustness against overfitting and can handle high-dimensional data.
- It can handle both numerical and categorical features, making it suitable for a wide range of problems.

Ensemble models combine multiple individual models to make predictions. Random Forest is a popular ensemble algorithm that constructs a multitude of decision trees and combines their outputs to make the final prediction. It performs well in handling high-dimensional datasets and provides feature importance rankings.

Boosting Model: Gradient Boosting Machines (GBM)

- GBM is a boosting algorithm that sequentially builds an ensemble of weak learners.
- It trains each model to correct the mistakes made by the previous models in the ensemble.
- GBM can capture complex relationships in the data and handle both numerical and categorical features.
- It typically uses decision trees as weak learners but can also use other models.
- GBM has gained popularity due to its high predictive power and ability to handle diverse datasets.

Boosting algorithms are another family of ensemble methods that iteratively train weak classifiers to create a strong classifier. Gradient Boosting Machines (GBM) is a widely used boosting algorithm. It builds the model in a stage-wise manner, optimizing the loss function by adding weak models to correct the mistakes of previous models.

Stacking

- Ensemble learning technique that combines multiple models with a meta-model.
- It involves training several base models on the training data and then using their predictions as input to a higher-level model (meta-model).
- The meta-model learns to combine the predictions of the base models to make the final prediction.
- Stacking can capture diverse patterns in the data and potentially improve predictive performance.

Stacking is an advanced ensemble technique that combines multiple models by training a meta-model on their predictions. It learns how to best combine the predictions from different models to make the final prediction. Stacking can often improve the performance of individual models and is worth exploring if you have the time and resources.