Kiejstut Bunikiewicz

Professor Brahma

BSAN 6070 – Machine Learning

3/15/21

CA04 – Ensemble Methods: Question Responses

Q.3.1 (Random Forest) Write you observations about the Classifier’s behavior with respect to the number of estimators.

With respect to the number of estimators, the results for both AUC and accuracy rate tend to fluctuate. The initial value with the lowest number of estimators performs the third best in terms of accuracy and the best in terms of AUC. However, the accuracy for 250 estimators is the highest. As the number of estimators increases after the peak of performance, the performance in both terms of AUC and accuracy steadily decreases.

Q.3.2 (Random Forest) Is there an optimal value of the estimator within the given range?

Within the given range, the estimator’s optimal value is 250 for the Random Forest Ensemble Model. This estimation is based on the 250 estimators having the highest accuracy rate of the range of estimators provided.

Q. 4. 1 (AdaBoost) Write you observations about the Classifier’s behavior with respect to the number of estimators.

For AdaBoost, the model’s performance in terms of accuracy and AUC decreases as the number of estimators increases. From 100 estimators to 500 estimators the performance is the same.

Q.4.2 (AdaBoost) Is there an optimal value of the estimator within the given range?

Within the given range, the optimal value for AdaBoost is 50 estimators in terms of AUC and accuracy.

Q. 4. 1 (Gradient Boost) Write you observations about the Classifier’s behavior with respect to the number of estimators.

The Gradient Boost classifier’s performance fluctuates based on the number of estimators. Generally, performance improves as the number of estimators increases. The performance values decrease slightly after reaching their maximum at 150 or 200, depending on accuracy or AUC. However, the higher number of estimators after the optimal value have much better performance than the estimators that are lower than the optimal values.

Q.4.2 (Gradient Boost) Is there an optimal value of the estimator within the given range?

Within the given range, the optimal value is 150 estimators if using accuracy and 200 estimators if using AUC as the determining metric.

Q. 4. 1 (Extreme Gradient Boost) Write you observations about the Classifier’s behavior with respect to the number of estimators.

Generally, performance improves as the number of estimators increases for the XGB model. The greatest performance boost came when the number of estimators increased from 100 to 150. Although performance dipped slightly afterwards, the model still performed much better with a higher number of estimators. The model peaked in performance in the 400 – 500 estimator range.

Q.4.2 (Extreme Gradient Boost) Is there an optimal value of the estimator within the given range?

In terms of AUC, 400 estimators is the optimal value. In terms of accuracy rate, 450 or 500 are the optimal values, with 450 estimators having a higher AUC than 500.