COMP8118-A6

Wednesday, October 12, 2022 6:13 PM

Q1. Bagging vs. Boosting

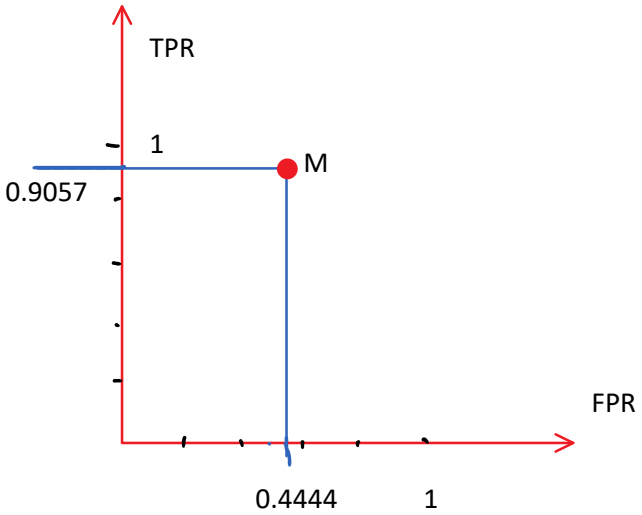
	Bagging	Boosting
Assumption	The weights are equal. Models built independently. Datasets are drawn randomly. Merges similar predictions. Decreases variance and fixes overfitting. Useful when the dataset is unstable.	The weights are different based on performance. New models are drawn from the previous models. Datasets are the previous misclassified data. Merges different predictions. Decreases bias and fixes underfitting. Useful when dataset is stable.
Construction process	Take repeated bootstrap samples from the training set by drawing some training samples randomly with replacements. The process includes: 1. Bootstrapping: Create k bootstrap samples. 2. Parallel Training: Train each sample individually in a classifier.	All data have equal weights first. The error is calculated. Then each misclassified sample's weight increases and the weights of the correct samples decrease.
Final aggregation of classifiers	3. Aggregation: Average all the classifiers' results. depending on the task (i.e. regression or classification), an average or a majority of the predictions are taken to compute a more accurate estimate. In the case of regression, an average is taken of all the outputs predicted by the individual classifiers; this is known as soft voting. For classification problems, the class with the highest majority of votes is accepted; this is known as hard voting or majority voting.	Combine (sum) all classifiers with some weight after t iterations.

Q2.

	t	f
+	48	12
-	15	5

a) Precision and Recall:
   i. P = true positive / (true positive + false positive) = 48 / 60 = 0.8
   ii. R = true positive / (true positive + false negative) = 48 / (48 + 5) = 0.9057

b) ROC
   i. TPR = true positive / (true positive + false negative) = 48 / (48 + 5) = 0.9057
   ii. FPR = false positive / (false positive + true negative) = 12 / (12 + 15) = 0.4444



Refrences:

1. Venugopal, Deepak, "Machine Learning, Ensemble Methods slides"
2. IBM, "Bagging", https://www.ibm.com/cloud/learn/bagging
3. Upgrad, "Bagging vs Boosting", https://www.upgrad.com/blog/bagging-vs-boosting/#:~:text=Bagging%20and%20Boosting%3A%20Differences,-As%20we%20said&text=Bagging%20is%20a%20method%20of,Boosting%20decreases%20bias%2C%20not%20variance.