Saagar Sanghavi, Jin Seok Won, Kevin Zhu

Multiplicative Weights and Boosting Quiz Questions

1. Bagging gets a better variance reducing effect when each of the models have lower or higher correlation?
    a. Lower correlation. When each of the models are perfectly correlated, there is no benefit from bagging.
2. What kind of models are considered weak learners?
    a. Any model that has accuracy of > 50%
3. What is the upper bound on regret for the multiplicative weights algorithm?
    a. $R_T \leq 2\sqrt{T \ln n}$
4. What is one way OMP can be related to boosting?
    a. Both make greedy choices that gets added to the overall predictor
    b. Keeps track of residual prediction errors
5. Adaboost is a specific method for which type of task?
    a. Binary classification
6. What is the name of the technique that allows for the generalization of boosting technique to arbitrary differentiable loss functions?
    a. Gradient boosting
7. What is an advantage that boosting takes advantage of but not bagging?
    a. Boosting updates the weights after each model is added to the collective model, making new greedy choices in the next step. There is no concept of order of models in Bagging and all models are used and weighted in the same way.
8. What is an advantage of using probabilistic allocation instead of discreet one?

    a. By using probabilistic models, we can use all of the probabilistic knowledge gained from the previous days, making an optimal bet on the next day. Choosing the highest probability expert all the time is more vulnerable to adversary.

9. How are decision boundaries of Boosting with linear weak classifiers differ from boundaries of decision trees?
    a. Boosting with linear classifiers can continue iterating even when all points are identified correctly. Decision trees cannot continue when there is no more points since it will hit its base case.
10. What is the advantage of using random forests over simple decision trees?
    a. Random forests are more stable than decision trees, having more smooth decision boundaries and more consistent output. It can also be parallelized unlike boosting algorithms.