**Quiz 4 Ensemble Learning: Stacking**

Study this brief and answer - why ensemble methods are able to out-perform any single classifier within the ensemble? Max 1 – 2 pages answer. <https://web.engr.oregonstate.edu/~tgd/publications/mcs-ensembles.pdf>

This paper doesn't discuss Stacking. Please explain why stacking helps and how is it different from the methods this paper discussed. Are the same 3 reasons applicable here?

There are 3 fundamental reason why ensemble methods out-perform single classifiers in general:

**The first reason is statistical.** This is related to the amount of data available for training. Given a learning algorithm that is searching a space H of Hypotheses for the best hypotheses, the statistical problem appears when the training data way smaller than the noted H space (Hypotheses space). One way to solve this problem is by constructing a model using different learning algorithms that could find different hypotheses in such space H and can provide good accuracy. On performing this ensemble of “accurate classifiers” (accurate classifier in this context refers to a classifier that can provide a better error rate than random guessing), the algorithm can average their votes and by doing this minimize the risk of selecting a wrong classifier/model. In other words by averaging the vote of all the models (accurate classifiers) we are after a good approximation of the true hypothesis (only known to nature).

**The second reason is computational.** This problem is associated with the inability of many learning algorithms to inherently find the global optima. Many of this algorithms get stuck in a local optima (given their inherent ability to perform a local search only), even when enough training data is provided (and therefore the statistical problem might not be present). One solution for this problem is to construct an ensemble model where the learning algorithms run the local search from many different points and by doing this the final ensemble might improve by obtaining a better approximation of the true hypothesis as we remove the constrain of getting stuck in the same local optima.

**The third reason is representational.** This shortcoming of standard learning algorithms relates to the inability of any individual hypotheses available on the H space to truly represent the true function f that can best represent the problem at hand. By creating weighted sums of different hypotheses coming from the space H it might be possible to get closer to the actual true function f.

Ensemble models provide a way to minimize these 3 problematic areas for standard learning algorithms and therefore attempt to find a better approximation of such true unknown function for the problem at hand.

In this paper it was mostly discussed ensemble methods from the perspective of bagging and boosting, but not stacking. Stacking is an ensemble machine learning algorithm that uses a meta-learning to learn how to best combine the predictions from two or more ML algorithms. The main difference with respect to other ensemble models is that instead of using trivial functions (as majority voting for classifiers or averaging for regressions) to aggregate the predictions, stacking trains a model (blender) to perform this process.

The 3 reasons on why ensemble models perform better than individual models still apply for stacking (stacked generalization) as the model looks a way to overcome the statistical, computational and representational problems that individual learning algorithms present, but instead of overcoming the statistical problem by averaging or majority voting the single prediction to obtain a final better prediction it uses a meta model to perform such aggregation as mentioned before.