Question 1

Suppose we are given a train set and test set, that came from the same distribution. We want to use stacking and choose between two validation schemes described in the reading material.

Select the true statements about validation schemes.

Correct answers:

- Scheme e) gives the validation score with the least variance, if compared to schemes a) -- d).
- Scheme d) is less efficient from computational perspective than scheme a). That is, if a dataset is very large, scheme a) is usually preferred over scheme d).

Incorrect answers:

- Scheme d) is more efficient from computational perspective than scheme a). That is, if a dataset is very large, scheme d) is usually preferred over scheme a).

Question 2

Definition: we will call a validation scheme fair if the set, that we use to validate meta-models comes from the same distribution as the meta-test set. In other cases we will call validation scheme leaky. In other words in a fair validation scheme the set that we use to validate meta-models was not used in any way during training first level models.

Select fair validation schemes. The definition for the schemes can be found in the reading material.

Correct answers:

a) Simple holdout scheme

d) Holdout scheme with OOF meta-features

e) KFold scheme with OOF meta-features

Incorrect answers:

b) Meta holdout scheme with OOF meta-features

c) Meta KFold scheme with OOF meta-features

Question 3

Which of the following ensembling methods can potentially learn "conditional averaging" (video 1)?

Correct answers:

- Boosting on trees
- Stacking

Incorrect answers:

- Weighted average
- Bagging

Question 4

The benefits of the weighted average compared to more advanced ensembling techniques is that

Correct answers:

- It is less prone to overfitting. Yes! A very small number of parameters rarely lead to overfitting.
- It is faster to implement and to run. Yes! It is much easier to implement than stacking.

Incorrect answers:

- It usually gives better quality. Well, in usually stacking can do better.

Question 5

In general case, which set of base models is probably the best for stacking?

Correct answers:

- [SVM, GBDT, Neural Network, kNN]. This set contains models from 4 different classes, so it is the most diverse set and it should be in general case be the best for stacking

Incorrect answers:

- [Random Forest, Extra Trees Classifier, GBDT, RGF]. All this are tree-based model, this set is not diverse enough
- [kNN, SVM, Logistic Regression, Neural Net]. SVM and Logistic Regression are linear models, so this set contains only 3 different classes of models (linear, kNN, NN)
- [Logistic Regression, SVM, Random Forest, Extra Trees Classifier, GBDT]. This set contain only models from 2 families (linear and tree-based) so it is not the best choice

Question 6

Suppose we are given a classification task. In a simple two model linear mix we usually use weights α for the first model and β for the second one. The coefficients are usually chosen such that α+β=1, because convex combination of probability vectors is a probability vector.Still, sometimes it is beneficial to tune α and beta independently, e.g. mix with α=0.1 and β=0.8 works best.

However, for some metrics it never makes sense to tune α and β independently. That is, searching for independent α and β will never give you better results than searching for weights, constrained to be β=1−α. Select such metrics.

Correct answers:

- AUC. AUC is only sensitive to the order of the objects, so AUC is the same for (1) predictions (2) same predictions multiplied by a positive constant. Then, for any α, β, dividing the predictions by $\alpha + \beta$ will not change AUC but make mixing coefficients sum up to one. So it does not make sense to explore α, β independently.
- Accuracy (implemented with argmax) It follows from the fact, that similarly to AUC, argmax position will not change if all the predictions multiplied by a constant.

Incorrect answers:

- Hinge loss
- LogLoss

✓ Complete   Go to next item