

Testing/Validation and Ensemble Methods Quiz

NAMES:

1. Training/Validation/Testing

- (a) You and your friend are arguing over whether to use 3-fold or 20-fold cross-validation to tune the hyperparameters of your learning algorithm. What arguments might be made for the two alternatives?

- (b) As an analyst at a car company, you are attempting to build a machine learning model that will determine whether or not a driver is too impaired to drive. Your training set contains 20 attributes describing driver state and behavior (swerviness, blink rate etc.) Along with a class label describing whether or not a collision occurred during the trip. Your training set consists of 2000 collisions and 2000 non-collisions. After training a random forest, your test set accuracy is 89%. Your supervisor asks whether you could get better results with a larger training set. How might you answer her question using only the labeled data that you currently have?

2. Random Forests

- (a) Bagging is an ensemble method that involves training multiple base classifiers on different subsets of the original training data, then allowing those classifiers to vote. How is the Random Forest algorithm different from just applying bagging to our standard decision tree algorithm? Based on your answer, do you think that the Random Forest improvement is likely be helpful for one-dimensional training data?

- (b) Random forests generally perform better than individual decision trees. Can you imagine a situation where a simple decision tree may be preferable to a random forest?

- (c) Random forests are typically composed of fully-expanded decision trees. In contrast, when using a simple decision tree, we are likely to use prune the final tree or to carefully tune the tree depth as a hyperparameter. How do you explain the difference?

- (d) The scikit-learn implementations of both the `DecisionTreeClassifier` and the `RandomForestClassifier` classes have an attribute named `feature_importances_`. The documentation for this attribute just says “Return the feature importances (the higher, the more important the feature).” How might you calculate feature importance from a decision tree? How could your approach be extended to a random forest? How might knowing “feature importance” be useful?

- (e) Why are Extremely Random Trees typically much faster to train than Random Forests?

3. AdaBoost

The AdaBoost algorithm can be summarized as follows:

During training, repeat the following for N iterations:

- Train a base C_j classifier on the weighted data set (initially, all weights are equal).
- Apply the classifier to all m training examples and calculate the weighted error (sum of the weights of the misclassified samples):

$$r_j = \sum_{\substack{i=1 \\ \hat{y}_j^{(i)} \neq y_j^{(i)}}}^m w_i$$

- Calculate the weight/importance of the new classifier:

$$\alpha_j = \ln \frac{1 - r_j}{r_j}$$

- Update the sample weights according to:

$$w_i^{(j+1)} = \begin{cases} w_i^{(j)} & \text{if } C_j(x_i) = y_i \\ e^{\alpha_j} w_i^{(j)} & \text{if } C_j(x_i) \neq y_i \end{cases}$$

- Normalize the weights so that they sum to 1:

$$w_i^{(j+1)} \leftarrow \frac{w_i^{(j+1)}}{Z_{j+1}}$$

where Z_j is the sum of the pre-normalized weights: $Z_{j+1} = \sum_{i=1}^m w_i^{(j+1)}$

At prediction, time select the class that gets the most (weighted) votes:

$$C^*(x) = \arg \max_k \sum_{\substack{j=1 \\ C_j(x)=k}}^N \alpha_j$$

Consider the following training set and initial weights:

x :	.1	.3	.6	.9
y :	-1	1	1	-1
$\mathbf{w}_0$:	.25	.25	.25	.25

- (a) Assume that C_0 has been created, with the following result:

x :	.1	.3	.6	.9
$\hat{y}_0$:	1	1	1	-1
correct?	N	Y	Y	Y

What is r_0 ? What is α_0 ? Show your work.

(b) Fill in the table below with the updated weights. Show your work.

$x:$	.1	.3	.6	.9
$\mathbf{w}_1:$				

(c) Now assume that C_1 performs as follows:

$x:$	.1	.3	.6	.9
$\hat{y}_1:$	-1	1	1	1
correct?	Y	Y	Y	N

Calculate r_1 , α_1 and the updated weights:

$x:$	.1	.3	.6	.9
$\mathbf{w}_2:$				

(d) Given that C_0 classifies a particular point as -1 and C_1 classifies it as 1, what class will be selected by the ensemble? Justify your answer.