Ensemble Methods

Quiz Questions

20 Multiple Choice Questions

Machine Learning Course

Question 1

Which of the following is NOT an ensemble method discussed in the slides?

A. Majority voting

B. Bagging

C. k-means clustering

D. Random forests

Question 2

Majority voting is guaranteed to reduce error (compared to an individual classifier) under which conditions?

A. Classifiers are independent and each has error rate > 0.5

B. Classifiers are independent and each has error rate < 0.5

C. Classifiers are identical and deterministic

D. There is only one classifier

Question 3

In the slides, the ensemble error of a majority vote classifier with n independent base classifiers, each with error ε, is written as:

A. ε = ε

B. ε = 1 - ε

C. ε = Σ (n choose k) ε (1-ε)

D. ε = ε / n

Question 4

Compared to hard voting, soft voting in the slides uses:

A. Only the class labels from each classifier

B. The class probabilities from each classifier

C. Randomly sampled labels from each classifier

D. The loss values from each classifier

Question 5

Three classifiers output $h_1(x) \rightarrow [0.9,0.1]$, $h_2(x) \rightarrow [0.8,0.2]$, $h_3(x) \rightarrow [0.4,0.6]$ with weights $w=[0.2,0.2,0.6]$. What is the predicted class?

A. Class 0 (because $p(j=0|x) = 0.58$)

B. Class 1 (because $p(j=1|x) = 0.72$)

C. Class 1 (because $p(j=1|x) = 0.42$)

D. Both classes are tied

Question 6

In bagging, bootstrap samples are drawn from a dataset of size m. Approximately what fraction of UNIQUE training examples appears in a given bootstrap sample?

A. 100%

B. About 50%

C. About 63.2%

D. About 36.8%

Question 7

Which type of base learner typically benefits the MOST from bagging?

A. Very simple linear models with low variance

B. Unpruned decision trees with high variance

C. k-means clustering

D. Pre-trained neural networks with fixed weights

Question 8

The main goal of bagging is to:

A. Reduce bias by using very simple models

B. Reduce variance by averaging over bootstrap models

C. Reduce training set size

D. Convert regression models into classifiers

Question 9

In the general boosting framework, which of the following steps is NOT part of the basic loop?

A. Initialize a uniform weight vector over training examples

B. Apply a weak learner to the weighted training data

C. Decrease weights of misclassified examples

D. Increase weights of misclassified examples

Question 10

In the slides, the weak learner used to illustrate AdaBoost is:

A. A deep neural network

B. A decision tree stump

C. A full random forest

D. A k-nearest neighbors classifier

Question 11

For squared error loss in gradient boosting, the pseudo-residuals turn out to be:

A. The true labels (y)

B. The predictions (h(x))

C. The negative of the residuals (-(y - h(x)))

D. The residuals (y - h(x))

Question 12

In gradient boosting, the learning rate α (0 < α ≤ 1) mainly controls:

A. The maximum depth of each tree

B. How strongly each new tree's predictions are added to the current model

C. The number of classes in the target

D. Whether we use classification or regression

Question 13

According to the slides, XGBoost is best described as:

A. A clustering algorithm based on decision trees

B. A scalable implementation of gradient boosting with extra regularization and optimization tricks

C. A simple bagging method with linear models

D. A method that only uses random subspaces, not trees

Question 14

LightGBM, as described in the slides, is characterized by:

A. Very slow training and high memory usage

B. Faster training, lower memory usage, and good accuracy on large-scale data

C. Being limited to binary classification

D. Not supporting parallel or GPU learning

Question 15

Random forests can be described as:

A. Gradient boosting with shrinkage

B. Bagging of decision trees plus random feature subsets

C. A single large decision tree with pruning

D. A stacking method with a neural network meta-learner

Question 16

Tin Kam Ho's random subspace method differs from Breiman's trademark random forest mainly because:

A. Ho's method uses boosting instead of bagging

B. Ho's method assigns a random feature subset to each TREE, while Breiman chooses random feature subsets at each NODE

C. Breiman's method does not use trees at all

D. Ho's method uses regression trees only

Question 17

A simple rule of thumb from the slides for the number of features to consider at each split in a random forest is:

A. $\sqrt{m}$, where m is the number of features

B. $m/2$

C. m (all features)

D. $\lfloor \log_2(m) + 1 \rfloor$

Question 18

In the generalization error bound PE $\leq \bar{\rho}(1-s^2)/s^2$, $\bar{\rho}$ and s represent:

A. Mean squared error and sample size

B. Average correlation among trees and strength of the ensemble

C. Learning rate and number of estimators

D. Bias and variance

Question 19

Compared to standard random forests, Extremely Randomized Trees (ExtraTrees) introduce ADDITIONAL randomness by:

A. Randomly choosing tree depth for each tree only

B. Randomly choosing both feature subsets and split thresholds, rather than optimizing splits

C. Removing bootstrap sampling

D. Using k-means to choose splits

Question 20

The naive stacking procedure (training base models and then a meta-classifier on their predictions on the SAME training set) has which main issue?

A. It cannot be used for regression

B. It is computationally impossible

C. It causes data leakage and overfitting of the meta-classifier

D. It requires deep neural networks