Intro to Big Data Science — Spring 2021-2022

Name: 刘乐奇 **ID No.:** 1201327

Quiz 4

To receive credit, this worksheet MUST be handed in at the end of the class.

B 1. Suppose you are trying to decide among a few different choices of kernel and are also choosing parameters such as C, σ^2 in the Gaussion kernel $\exp(-\frac{\|x_1-x_2\|^2}{2\sigma^2})$, etc. How should you make the choice?

(A) Choose whatever performs best on the training data.

(B) Choose whatever performs best on the cross-validation data.

(C) Choose whatever performs best on the test data.

(D) Choose whatever gives the largest SVM margin.

A 2. Which is incorrect about kernel?

(A) There is no need to have the explicit expression of the map in kernel methods.

(B) Kernel matrices are the semi-positive definite.

(C) Kernel functions map the feature into low dimensional space.

(D) Kernel functions lead to the linearly separable problems in the new space.

A 3. Which is incorrect about support vector machine (SVM)?

(A) The objective of SVM is to minimize the margin around the separating plane.

(B) The samples at the boundary of the margin are called support vectors.

(C) SVM can be solved in dual formulation.

(D) SVM can also be used to classify samples which are not linearly separable.

AB 4. In random forest, which of the following quantities are randomly selected? (Select all correct answers)

(A) Sample data

(B) Features

(C) Number of leaf nodes

(D) Number of trees

AD 5. Which of the following is incorrect? (Select all answers that apply)

(A) Bagging is a sequential training approach

(B) Bagging does resampling to generate many data sets

(C) Out-of-bag samples are the samples that are not selected in the sampling process, therefore they are useless.

(D) Bagging is easy to interpret

ABC 6. Which strategies could be used to combine the base models (classifiers or regressors) to form an ensemble model? (Select all correct answers)

(A) Majority vote

(B) Weighted average

(C) Average

(D) Random sampling

7. (AdaBoost) Consider building an ensemble of decision stumps G_m with the AdaBoost algorithm

$$f(x) = \text{sign}\Big(\sum_{m=1}^{M} \alpha_m G_m(x)\Big).$$

Figure 1 displays a few labeled point in two dimensions as well as the first stump we have chosen. A stump predicts binary ± 1 values, and depends only on one coordinate value (the split point). The little arrow in the figure is the normal to the stump decision boundary indicating the positive side where the stump predicts $+1$. All the points start with uniform weights.

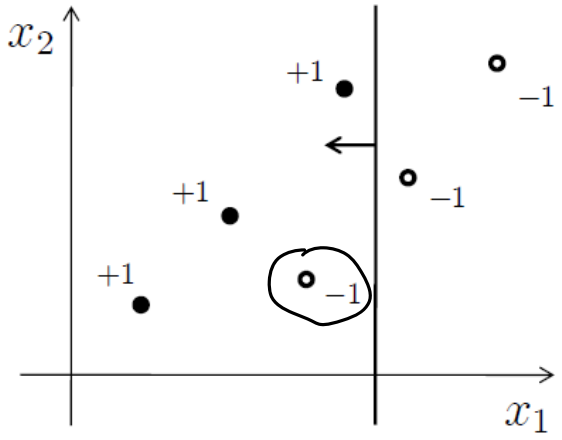


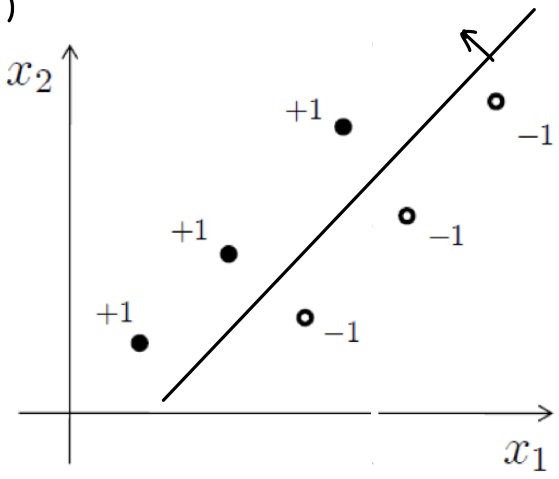
Figure 1: Labeled points and the first decision stump. The arrow points in the positive direction from the stump decision boundary.

(a) Circle all the point(s) in Figure 1 whose weight will increase as a result of incorporating the first stump (the weight update due to the first stump).

(b) Draw in the same figure a possible stump that we could select at the next boosting iteration. You need to draw both the decision boundary and its positive orientation.

(c) Will the second stump receive higher coefficient in the ensemble than the first? In other words, will $\alpha_2 > \alpha_1$? Briefly explain your answer. (no calculation should be necessary).

(b)



(c) Yes.

$$\epsilon_m = \frac{\sum_{i=1}^{n} w_i^{(m)} I(y_i \neq G(x_i))}{\sum_{i=1}^{n} w_i^{(m)}}$$

Since $I(y_i \neq G(x_i))$ decreases after second stump, $\epsilon_2 < \epsilon_1$

$$\alpha_m = \log\frac{1-\epsilon_m}{\epsilon_m} = \log\left(\frac{1}{\epsilon_m} - 1\right)$$

with smaller ϵ_m, $\alpha_2 > \alpha_1$