Machine Learning

V. Adamchik | CSCI 567 | Fall 2019

Discussion Set 8 | University of Southern California

Principal Components Analysis

Boosting

Problem 1

Given 3 data points in 2-d space, (1, 1), (2, 2) and (3, 3), what is the first principal component?

Problem 2

The trace Tr of a square matrix is defined as the sum of the elements on the main diagonal.

Prove that for the covariance matrix C

$$Tr[C] = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_n^2 = \lambda_1 + \lambda_2 + \ldots + \lambda_n$$

where λ_i and σ_i are eigenvalues and variances.

Problem 3

Find the second PCA value:

$$\max(v_2^T X^T X v_2),$$

subject to

$$v_2^T v_2 = 1$$

$$v_2^T v_1 = 0$$

Problem 4

Which of the following is wrong about PCA?

(A) PCA outputs a compressed dataset that is a linear transformation of the original dataset.

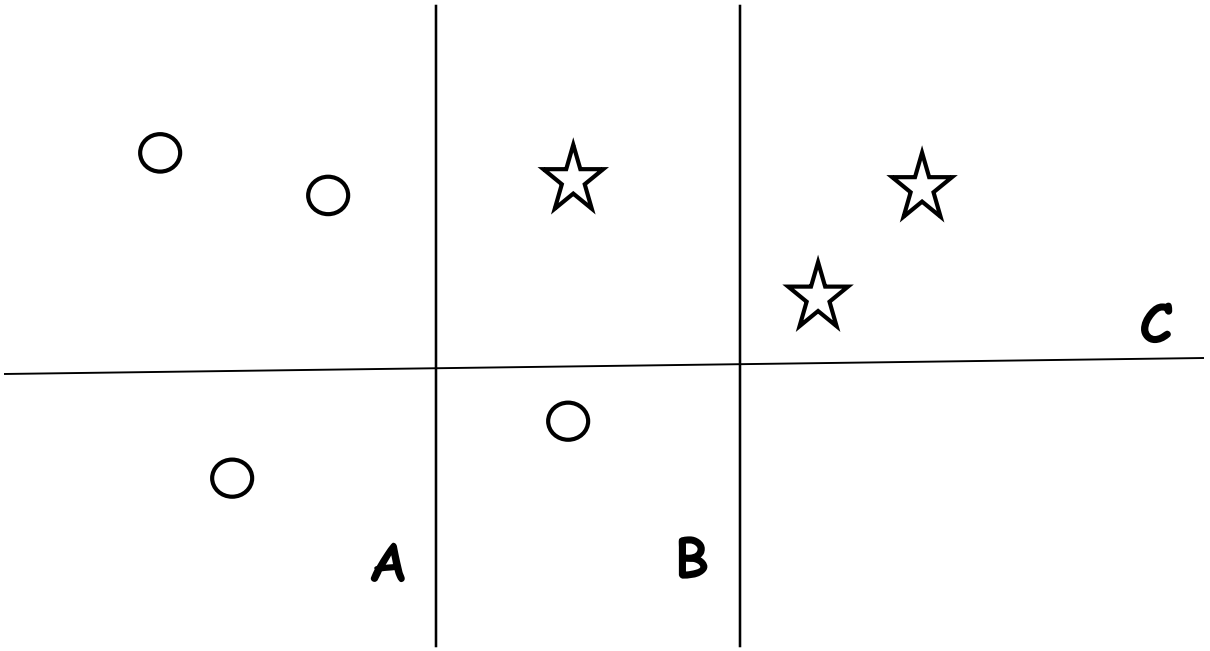
(B) The first principal component is the eigenvector of the covariance matrix with the largest eigenvalue.

(C) The first step of kernel PCA is to center the original dataset.

(D) Kernel PCA requires computing eigenvalues and eigenvectors of the Gram matrix.

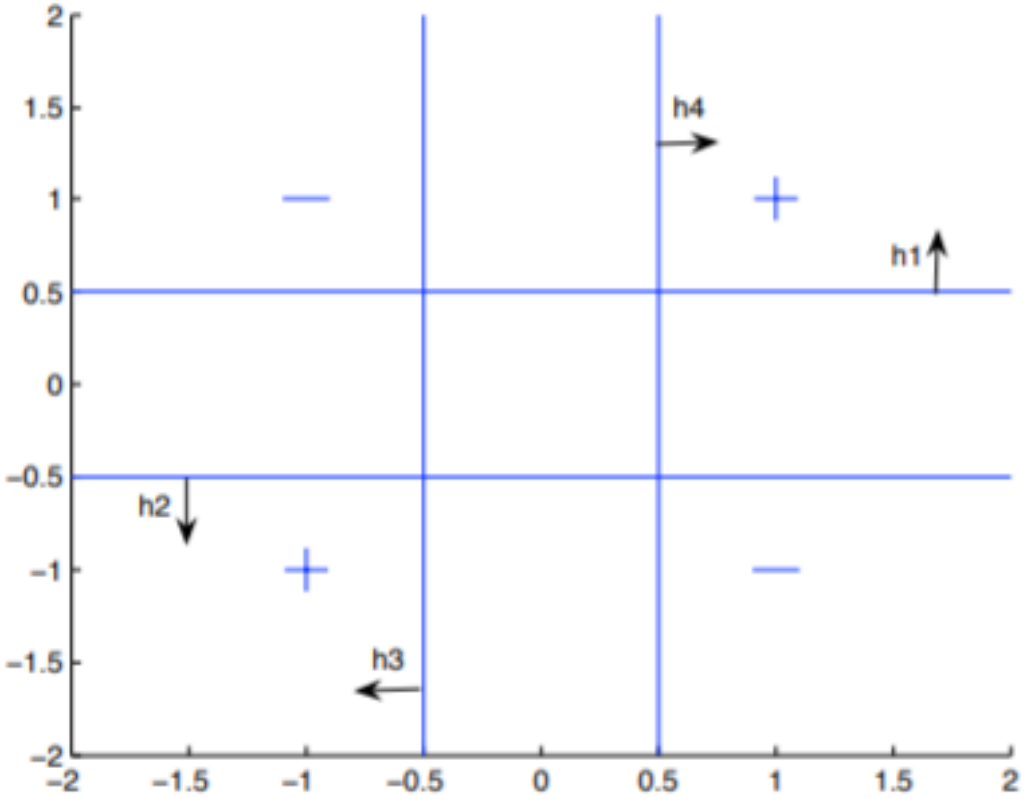
Problem 5

The diagram shows training data for a binary concept where positive examples are denoted by a star. Also shown are three decision stumps (A, B and C) each of which consists of a linear decision boundary. Suppose that AdaBoost chooses A as the first stump in an ensemble and it has to decide between B and C as the next stump. Which will it choose?



Problem 6

Consider the four binary classifiers below. The arrow means that the corresponding classifier classifies every data point in that direction as +. Prove that there are no weights $\beta_1, \ldots, \beta_4$, that make the ensemble $\Sigma\beta_i h_i$ classifier consistent with the data.



Problem 7

T/F. In AdaBoost weights of the misclassified examples go up by the same multiplicative factor.

T/F. A weak learner with less than 50% accuracy does not present any problem to the AdaBoost algorithm.

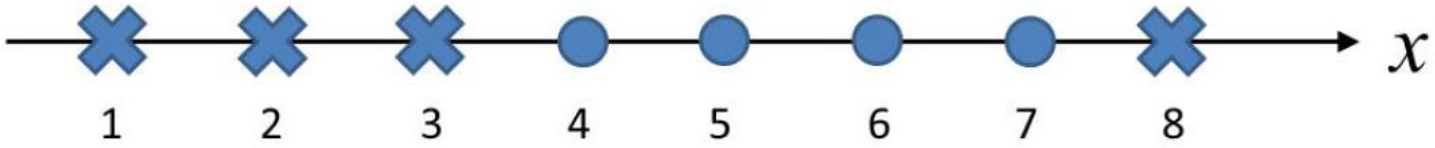
T/F. Once a weak classifier is picked in a particular round, it will never be chosen in any following rounds.

T/F. If you have 100 training points, the updated weight of a training point cannot be greater than $\frac{1}{2}$.

T/F. AdaBoost will definitely achieve zero training error regardless of the type of weak classifier it uses, provided enough rounds T are performed.

Problem 8

Imagine running AdaBoost with a 1-dimensional training set of 8 examples as shown



Circles mean y = +1 and crosses mean y = -1. The number under each example is its x coordinate. The base classier set H consists of all decision stumps such that

$$h_i(x) = \begin{cases} s, & \text{if } x > b \\ -s, & \text{otherwise} \end{cases}$$

1. Run AdaBoost for two rounds and compute β_1 and β_2.
2. Compute the training error of the final classifier H.