Question 1: Short Questions

(a) Suppose you want to apply AdaBoost algorithm on a dataset. You set N_1 samples for training and the rest for testing. Which of the following is true?

(i) The difference between training and testing error increases as number of training data N_1 increases

(ii) The difference between training and testing error decreases as number of training data N_1 increases

(iii) The difference between training and testing error will not change

(iv) None of the above

(b) Which of the following algorithm is not an example of ensemble learning algorithm?

(i) Random forest

(ii) Adaboost

(iii) Decision tree

(c) Which of following are valid covariance matrices?

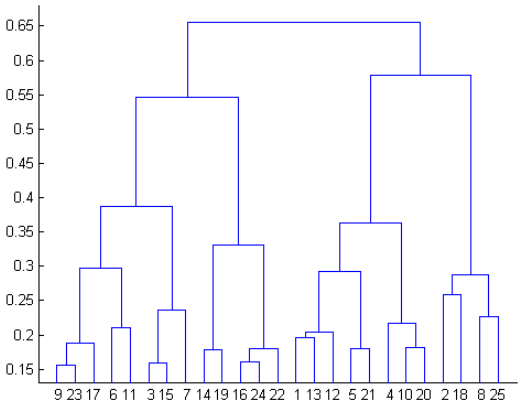
(i) $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$

(ii) $\mathbf{B} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$

(iii) $\mathbf{C} = \begin{bmatrix} 0 & 1 \\ 1 & 2 \end{bmatrix}$

(iv) $\mathbf{D} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}$

(d) After performing agglomerative hierarchical clustering on a dataset, you observed the following dendrogram. Which of the following conclusion(s) can be drawn from the dendrogram? One or more answers might be correct.



(i) There were 50 data samples in this problem.

(ii) The closest data samples in this dataset are samples 9 and 23.

(iii) The closest data samples in this dataset are samples 8 and 25.

(e) In terms of the bias-variance trade-off, which of the following is/are substantially more harmful to the test error than the training error?

(i) Bias

(ii) Variance

(iii) Loss

(iv) None of the above

Question 2: Principal Component Analysis

Consider 3 data points in the 2-d space: $\mathbf{x_1} = [-1, -1]^T$, $\mathbf{x_2} = [0, 0]^T$, $\mathbf{x_3} = [1, 1]^T$.

(a) What is the first principal component (write down the actual vector)?

(b) If we project the original data points into the 1-d subspace by the principal component you choose, what are their coordinates in the 1-d subspace? And what is the variance of the projected data?

(c) For the projected data you just obtained above, now if we represent them in the original 2-d space and consider them as the reconstruction of the original data points, what is the reconstruction error?