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

(ii) is the correct answer. As we have more and more data, training error increases and testing error decreases. And they all converge to the true error.

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

(i) Random forest

(ii) Adaboost

(iii) Decision tree

(iii) is the correct answer. Decision tree does not combine multiple classifiers.

(c) Which of following are valid covariance matrices?

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

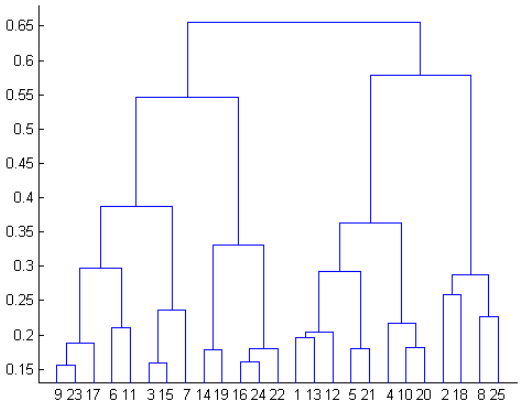
(ii) $\mathbf{B} = \begin{bmatrix} 1 & -1 \\ -1 & 2 \end{bmatrix}$

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

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

(ii) and (iv) are the correct answers. The covariance matrix of a Gaussian distribution should be symmetric. (iii) does not represent valid covariance, because the first element, corresponding to the variance of the first variable, is zero.

(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.

(ii) is the correct answer. The total number of samples is 25. The samples in the beginning of the dendogram are the ones first clustered together and are the closest, therefore 9 and 23 are the closest.

(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

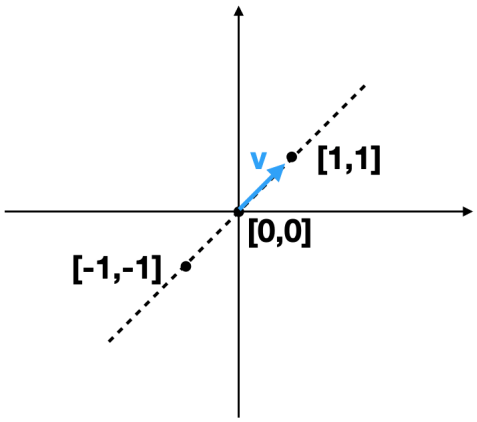
(ii) is the correct answer. If the model has a high variance, it is more likely to result in erroneous estimates on the test data, therefore can be substantially more harmful to the test error.

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)?

All points are located along the line $y = x$, therefore the first PCA dimension would be $\mathbf{v} = [1, 1]^T$. After normalizing $\mathbf{v}$ to have $l2$-norm equal to one, we get $\mathbf{v} = [\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}]^T = [\frac{\sqrt{2}}{2}, \frac{\sqrt{2}}{2}]^T$.



(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?

$z_1 = \mathbf{x_1}^T\mathbf{v} = -\sqrt{2}$
$z_2 = \mathbf{x_2}^T\mathbf{v} = 0$
$z_3 = \mathbf{x_3}^T\mathbf{v} = \sqrt{2}$

(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?

$\hat{\mathbf{x}}_\mathbf{1} = z_1\mathbf{v} = [-1, -1]^T$
$\hat{\mathbf{x}}_\mathbf{2} = z_2\mathbf{v} = [0, 0]^T$
$\hat{\mathbf{x}}_\mathbf{3} = z_3\mathbf{v} = [1, 1]^T$

All points are perfectly reconstructed (i.e., $\hat{\mathbf{x}}_\mathbf{n} = \mathbf{x_n}$), therefore the reconstruction error is zero.