Q1-1: Are these statements true or false?
(A) Learning lower dimensional representation can save memory usage.
(B) Learning lower dimensional representation can remove redundancies and noises in data.

1. True, True
2. True, False
3. False, True
4. False, False

Q1-1: Are these statements true or false?
(A) Learning lower dimensional representation can save memory usage.
(B) Learning lower dimensional representation can remove redundancies and noises in data.

1. True, True ←
2. True, False
3. False, True
4. False, False

(A) As is shown in the lecture.
(B) Actually one way to do reduce the size of a representation is to find and remove redundancies. Identifying and removing more redundancies allows the dimensionality reduction algorithms to achieve more compression while discarding less information. Usually noises will not lie in principal features, so can be filtered out.

Q1-2: Are these statements true or false?
(A) When we use PCA, we need data to be labelled.
(B) PCA extracts the variance structure from high dimensional data such that the variance of projected data is minimized.

1. True, True
2. True, False
3. False, True
4. False, False

Q1-2: Are these statements true or false?
(A) When we use PCA, we need data to be labelled.
(B) PCA extracts the variance structure from high dimensional data such that the variance of projected data is minimized.

1. True, True
2. True, False
3. False, True
4. False, False ←

(A) PCA is an unsupervised method, so it does not need labels of data.
(B) The variance of projected data is maximized, so projected data can contain most information of original data.

Q2-1: Are these statements true or false?
(A) The i-th principal component is taken as the direction that is orthogonal to the (i-1)-th principal component and maximizes the remaining variability.
(B) Different individual principal components (directions) are linearly uncorrelated.

1. True, True
2. True, False
3. False, True
4. False, False

Q2-1: Are these statements true or false?
(A) The i-th principal component is taken as the direction that is orthogonal to the (i-1)-th principal component and maximizes the remaining variability.
(B) Different individual principal components (directions) are linearly uncorrelated.

1. True, True
2. True, False
3. False, True ←
4. False, False

(A) The i-th principal component should be orthogonal to all the first (i-1) principal components.
(B) Since principal components are orthogonal to each other, different individual principal components are linearly uncorrelated.

Q2-2: Are these statements true or false?
(A) The principal component with the largest eigenvalue maximizes the reconstruction error.
(B) The dimension of original data representation is always higher than the dimension of transformed representation of PCA.

1. True, True
2. True, False
3. False, True
4. False, False

Q2-2: Are these statements true or false?
(A) The principal component with the largest eigenvalue maximizes the reconstruction error.
(B) The dimension of original data representation is always higher than the dimension of transformed representation of PCA.

1. True, True
2. True, False
3. False, True
4. False, False

(A) The principal component with the largest eigenvalue captures the maximum amount of variability which is equivalent to minimum reconstruction error.
(B) If the matrix XX^T is full-rank, they can be of the same dimension.

Q3-1: Are these statements true or false?
(A) Ignoring the components of small eigenvalues will not lose information.
(B) With better parameter tuning, we can get a better first principal component which maximizes the variability more precisely.

1. True, True
2. True, False
3. False, True
4. False, False

Q3-1: Are these statements true or false?
(A) Ignoring the components of small eigenvalues will not lose information.
(B) With better parameter tuning, we can get a better first principal component which maximizes the variability more precisely.

1. True, True
2. True, False
3. False, True
4. False, False ←

(A) We may lose some information if we ignore the components of small eigenvalues because there is still some variability at those directions. But they are of little significance due to small variability.
(B) The computation of PCA usually only involves eigenvalue decomposition, which includes no parameter tuning.

Q3-2: Assume that we have the following covariance matrix XX^T, please calculate the first principal component.

$$XX^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

1. $[1, 0, 0]^T$
2. $[0, 1, 0]^T$
3. $[0, 0, 1]^T$
4. $[1, 2, 3]^T$

Q3-2: Assume that we have the following covariance matrix XX^T, please calculate the first principal component.

$$XX^T = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{bmatrix}$$

1. $[1, 0, 0]^T$
2. $[0, 1, 0]^T$
3. $[0, 0, 1]^T$ ←
4. $[1, 2, 3]^T$

As we can see XX^T is already a diagonal matrix, and its largest eigenvalue is 3. The corresponding eigenvector is $e_3 = [0, 0, 1]^T$.