EECS545 Lecture 16 Quiz Solutions

1. What is the main goal of PCA?

   (a) To cluster the dataset in a latent subspace

   (b) To reduce the dimensionality of a dataset

   (c) To maximize the correlation between the features in a dataset

   (d) To minimize the variance in a dataset

   Solution: (b). PCA is a dimensionality reduction method.

2. Which of the following are true about PCA? Choose all that apply.

   (a) The principal component vectors discovered by PCA are always orthogonal to each other.

   (b) It is possible to kernerlize PCA algorithm.

   (c) PCA can be used for feature selection.

   (d) PCA is an supervised learning algorithm.

   (e) PCA requires an assumption that the data is normally (Gaussian) distributed.

   Solution: (a), (b), and (c).

   (a) The principal component vectors are eigenvectors, which are always orthogonal to each other; (c) PCA transforms the data to smaller dimension that retains most of the information as a linear combination of the data, so we can find which features are important for best describing the variance in a broader sense; (d) PCA is an unsupervised learning method; (e) We did not assume so when deriving PCA.

3. In PCA, to find the principal components, we try to maximize ___________:

   (a) the data likelihood

   (b) the variance of the data in the feature space

   (c) the variance of the data projected onto the principal components

   (d) the average norm of the data points projected onto the principal components

   (e) the approximation error of the data projected onto the principal components

   (f) the matrix norm of the data covariance matrix

Solution: (c). Please note that (e) is what PCA tries to minimize, and that (d) may also be true when data is zero-centered.

4. What kind of computation(s) are being done in PCA? Choose all that are correct.

   (a) $\max_{\mathbf{u}:\|\mathbf{u}\|_2=1} \frac{1}{N} \sum_n \mathbf{u}^\top \mathbf{x}^{(n)}$

   (b) $\max_{\mathbf{u}:\|\mathbf{u}\|_2=1} \frac{1}{N} \sum_n (\mathbf{u}^\top \mathbf{x}^{(n)} - \mathbf{u}^\top \overline{\mathbf{x}})^2$

   (c) $\max_{\mathbf{u}:\|\mathbf{u}\|_2=1} \frac{1}{N} \sum_n \|\mathbf{x}^{(n)} - (\mathbf{u}^\top \mathbf{x}^{(n)})\mathbf{u}\|^2$

   (d) $\min_{\mathbf{u}:\|\mathbf{u}\|_2=1} \frac{1}{N} \sum_n \|\mathbf{x}^{(n)} - (\mathbf{u}^\top \mathbf{x}^{(n)})\mathbf{u}\|^2$

   (e) Find the eigenvector(s) with the largest eigenvalue(s) of the data feature matrix

   (f) Find the eigenvector(s) with the largest eigenvalue(s) of the data covariance matrix

   (g) Find the eigenvector(s) with the smallest eigenvalue(s) of the data covariance matrix

   (h) Find the eigenvector(s) with the smallest eigenvalue(s) of the data covariance matrix

Solution: (b), (d), and (f).

(a) Note that $\overline{\mathbf{x}} \neq 0$ in general. (b) This is the variance maximization objective we've seen in the lecture. (d) This is the minimum distortion objective which is equivalent to the maximization objective; note that $(\mathbf{u}^\top \mathbf{x}^{(n)})\mathbf{u}$ is the projection of $\mathbf{x}^{(n)}$ onto $\mathbf{u}$.

5. In PCA, let's suppose we are finding the first principal component $\mathbf{u}_1$ and in order to do so you have found the (real) eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_R$ such that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_R$, and their corresponding eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_R$ of the data covariance matrix. Then, how can you recover the projection variance for the first principal component?

Solution: λ_1.

Note that the variance of the first principal component is same as $\mathbf{u}_1^\top \mathbf{S} \mathbf{u}_1$. This equals λ_1 because $\mathbf{S}\mathbf{u}_1 = \lambda \mathbf{u}_1$.