Review quiz

Started: Oct 30 at 7:21pm

Quiz Instructions

Question 1 — 5 pts

Which of the following is the correct statement about scatterplot matrix?

- ○ Its main weakness is that we cannot use density estimation techniques.
- ◉ Its main weakness is that the number of plots increases quickly as we have more features (dimensions).
- ○ Scatterplot matrix should be avoided because it does not work for high dimensions.

Question 2 — 5 pts

What is the key assumption of dimensionality reduction process?

- ◉ The data distribution can usually be approximated using a low-dimensional subspace.
- ○ We can always find a good low-dimensional representation of the dataset, regardless of how the data is distributed.
- ○ The data is more or less uniformly distributed across the whole space.

Question 3 — 5 pts

Which is the correct explanation of the first principal component in PCA?

- ◉ The first PC is the axis of the largest variance in the data.

- ◯ The first PC may not be orthogonal to some of the subsequent PCs.
- ◯ The first PC is the axis of the smallest variance in the data.

Question 4

5 pts

Given the following data points, calculate its covariance matrix's eigenvector that corresponds to the first principal component. Write down the first coordinate of the component. For example if the first principal component is [x, y] , return x

(0, 2)

(2, 6)

(1, 4)

(2, 5)

(-1, -1)

You can calculate them by hand, but numpy also provides functions for calculating the covariance matrix and eigenvector & eigenvalues.

https://docs.scipy.org/doc/numpy-1.13.0/reference/generated/numpy.cov.html (https://docs.scipy.org/doc/numpy-1.13.0/reference/generated/numpy.cov.html)

https://docs.scipy.org/doc/numpy-1.13.0/reference/generated/numpy.linalg.eig.html (https://docs.scipy.org/doc/numpy-1.13.0/reference/generated/numpy.linalg.eig.html)

-0.421

Quiz saved at 7:25pm

Submit Quiz