CS251 HW 6 | Mon Apr 1, 2019 | Week 9

Name:

Question 1

a) Plug in and simplify the multivariate Gaussian PDF centered on the point $(2, 4)$, where $\vec{x} = [x, y]$ $\sigma_x^2 = 2$, $\sigma_y^2 = 4$, x is independent with respect to y. *Note:* there's no need to memorize this equation.

$$f(\vec{x}|\vec{\mu}, \Sigma) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} e^{-\frac{1}{2}(\vec{x}-\vec{\mu})^T \Sigma^{-1} (\vec{x}-\vec{\mu})}$$

b) Sketch the qualitative shape of the PDF.

c) A spectral decomposition of the covariance matrix from (a) yields the following eigenvalues ($\vec{e}$) and eigenvectors (column vectors in V). Assume that the i^{th} eigenvalue corresponds to the i^{th} eigenvector. What are the first (PC_0) and second (PC_1) principal component directions?

$$\vec{e} = [2, 4]$$

$$V = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

d) Superimpose arrows corresponding to the PC_0 and PC_1 directions, proportional to the amount of variance they account for, onto your diagram from (b).

e) Say that we dropped PC_1 before projecting samples from the multivariate Gaussian PDF (data) into PCA space. Sketch the qualitative shape of the transformed data. Label your axes. *Note:* This problem is not asking you to derive numeric values of the transformed coordinates.

Question 2

Create qualitative sketches of data consistent with the following covariance matrices:

a) $\begin{bmatrix} 10 & 5 \\ 5 & 2 \end{bmatrix}$

b) $\begin{bmatrix} 10 & -5 \\ -5 & 2 \end{bmatrix}$

c) $\begin{bmatrix} 5 & 0.1 \\ 0.1 & 2 \end{bmatrix}$

Question 3

Given the following data matrix in the usual format (*rows*: data points, *cols*: features/variables)

$$X = \begin{bmatrix} 4.17 & 7.2 & 0 & 3.02 & 1.47 \\ 0.92 & 1.86 & 3.46 & 3.97 & 5.39 \\ 4.19 & 6.85 & 2.04 & 8.78 & 0.27 \\ 6.7 & 4.17 & 5.59 & 1.4 & 1.98 \\ 8.01 & 9.68 & 3.13 & 6.92 & 8.76 \\ 8.95 & 0.85 & 0.39 & 1.7 & 8.78 \\ 0.98 & 4.21 & 9.58 & 5.33 & 6.92 \end{bmatrix}$$

a singular value decomposition yields:

$$U = \begin{bmatrix} -0.16 & -0.4 & -0.32 & -0.04 & 0.67 & 0.41 & 0.3 \\ -0.04 & 0.36 & -0.26 & -0.52 & 0.14 & -0.51 & 0.5 \\ -0.51 & -0.37 & -0.03 & -0.22 & -0.64 & 0.18 & 0.35 \\ 0.03 & 0.04 & -0.32 & 0.8 & -0.13 & -0.23 & 0.44 \\ 0.13 & -0.3 & 0.79 & 0.06 & 0.18 & -0.22 & 0.43 \\ 0.79 & -0.04 & -0.13 & -0.18 & -0.28 & 0.39 & 0.3 \\ -0.25 & 0.7 & 0.28 & 0.1 & 0.04 & 0.54 & 0.27 \end{bmatrix}$$

$$S = \begin{bmatrix} 10.03 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 9.77 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8.34 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 5.36 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3.43 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$V^T = \begin{bmatrix} 0.53 & -0.37 & -0.26 & -0.41 & 0.59 \\ -0.47 & -0.46 & 0.66 & -0.14 & 0.33 \\ 0.18 & 0.52 & 0.28 & 0.48 & 0.62 \\ 0.51 & 0.26 & 0.62 & -0.44 & -0.32 \\ -0.45 & 0.56 & -0.18 & -0.63 & 0.23 \end{bmatrix}$$

a) Identify (or recover, if necessary) the eigenvectors and eigenvalues.

b) Calculate the proportion of variance accounted for by the each of the k principal components, where $k = 1, 2, \ldots, 5$.

c) Calculate the *cumulative* proportion of variance accounted for by the top k principal components, where $k = 1, 2, \ldots, 5$.

d) If we wanted to reduce the dimensionality of our data while retaining 95% of the variance, which principal components would we drop?

Question 4

In this question, you will write a short Python/Numpy script to explore the connection between covariance matrices and singular value decomposition.

Consider the following data in the usual format (*rows*: data points, *cols*: features/variables).

$$X = \begin{bmatrix} 0 & -4 & 4 \\ -4 & 7 & -3 \\ 4 & -3 & -1 \end{bmatrix}$$

a) Follow these steps to obtain the eigenvectors/values of the covariance matrix and singular value decomposition.

1. The data are already centered, do not re-center them.
2. Compute the covariance matrix using the formula $X^T X/(N-1)$.
3. Use `np.linalg.eig` to compute the eigenvalues ($\vec{a}$) and eigenvectors (B).
4. Compute the singular value decomposition on the centered data (via `np.linalg.svd`).

b) Are the rows or columns of B the eigenvectors? How about V^T?

c) Compare your V^T (from SVD) to your eigenvector matrix B (from covariance). Do you notice anything odd? If so, why is this ok?

d) Print out U and V^T. Do you observe a relationship?

e) Change the entry (row 2, col 3) of X to a number not already in the matrix (e.g. 8). Does the relationship between U and V^T still hold?

f) Change the entry at (2, 3) back, then change the entry (row 3, col 3) to a number not already in the matrix (e.g. 8). Does the relationship between U and V^T still hold?

g) Add a row to X (another data point) with numbers not already in the matrix. Does the relationship between U and V^T still hold?

h) Based on your results from (e-g), what property do you think X should have for the relationship between U and V^T to hold?