CSE 3380 – Homework #9

Assigned: Thursday, April 12, 2018
Due: Thursday, April 19, 2018 at the end of class

Note the following about the homework:

1. You must show your work to receive credit.
2. If your submission has more than one page, staple the pages. If I have to staple it, the cost is 10 points.

Assignment:
Process

1. We have matrices
$$A = \begin{bmatrix} 18 & -4 \\ -4 & 12 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 15 & -4 \\ -4 & 15 \end{bmatrix}$$
Find the eigenvalues and eigenvectors of A. Then use the determinant and trace of both matrices to show with certainty that the eigenvalues of the matrices are not completely identical or that they could be identical. Do this without actually finding the eigenvalues of B.
2. We have matrices
$$A = \begin{bmatrix} 8 & 0 & -5 \\ 9 & -1 & -5 \\ 14 & -4 & -7 \end{bmatrix} \quad \text{and} \quad B = \begin{bmatrix} 2 & -4 & 1 \\ -1 & -1 & 1 \\ 4 & -4 & -1 \end{bmatrix}$$
Find the eigenvalues and eigenvectors of A. Then use the determinant and trace of both matrices to show with certainty that the eigenvalues of the matrices are not completely identical or that they could be identical. Do this without actually finding the eigenvalues of B. **Hint**: One of the eigenvalues of A is -1.

Applications

3. **(CS application: Principal Component Analysis)** (MATLAB solution) There are many ways to analyze data. One tool for data analysis is Principal Component Analysis (PCA), which was discovered over 100 years ago. PCA uses eigenvectors to determine the directions of maximum variance in a set of data; these eigenvectors are the principal components. One area of Computer Science in which PCA is used is computer vision.

   (a) The steps for producing the principal components of an $n \times 2$ matrix A in which each column is the x-y coordinates of a single data point is:

      i. Find the sample mean, μ. For our data this means a vector consisting of the average x value and the average y value.
      ii. Subtract the mean from each point
$$s_i = p_i - \mu$$
to produce a new set of data. These new values produce a matrix S with n rows. Look into the idea of *broadcasting* in numPy.

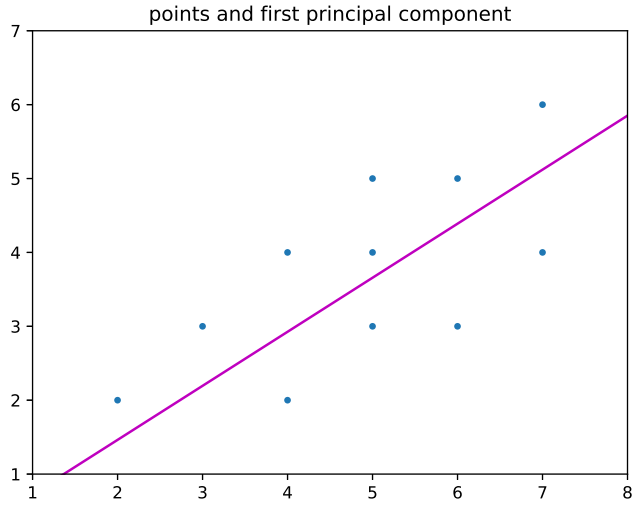

Figure 1: Data points with line showing direction of maximum variance.

iii. Produce $S^T S$. In statistical terms, this is the covariance matrix.

iv. Find the eigenvalues and eigenvectors of $S^T S$. If the eigenvalues are sorted from largest to smallest in absolute terms, i.e.,

$$|\lambda_1| \geq |\lambda_2| \geq \cdots$$

then the eigenvector $\vec{v}_1$ corresponding to λ_1 is the first principal component and points in the direction of the most variance in the data, the eigenvector $\vec{v}_2$ corresponding to λ_2 is the second principal component and points in the direction of the second most variance in the data and is also orthogonal to $\vec{v}_1$, and so forth.

(b) On the course website is a file, `pcaMain.py`, that contains an $n \times 2$ matrix (in the code it's written to look like $2 \times n$, but then I transpose), where each row is the x-y coordinates of one of the n points. This matrix is given to a function with the signature

```
shiftedPoints, covMatrix, fpc = pcaStudent(points)
```

where

- `shiftedPoints` is an $n \times 2$ matrix of the mean-shifted points
- `covMatrix` is a 2×2 matrix that is the covariance matrix of the mean-shifted points
- `fpc` is a 2×1 column vector of the first principal component of the covariance matrix

Write the function `pcaStudent()`, which will be placed in a file called `pcaStudent.py`. Your function should not be hard-coded to this particular data; use what is passed in. This includes not hard-coding to the number of points.

Note that you can learn more about PCA at [Ath13] and [Vek13].

General requirements about the Python problems:

a) As a comment in your source code, include your name.

b) The Python program should do the work. Don't perform the calculations and then hard-code the values in the code or look at the data and hard-code to this data unless instructed to do so.

c) Your function should use the data passed to it and should work if I were to change the data.

d) The program should not prompt the user for values or read from files unless instructed to do so.

e) I don't want your function to cause anything to print; leave it to the `XXXmain.py` file to do the printing of what is returned.

To submit the Python portion, do the following:

a) Create a directory using your net ID in lowercase characters. This should be something of the form `abc1234`.

b) Place your .py files in this directory.

c) Zip the directory, not just the files within the directory. You must use the zip format and the name of the file, assuming your net ID is `abc1234`, will be `abc1234.zip`.

d) Upload the zip'd file to Blackboard.

References

[Ath13] Vassilis Athitsos. CSE 6367 Computer Vision: notes for Principal Component Analysis. http://vlm1.uta.edu/~athitsos/courses/old/cse6367_spring2012/lectures/11_pca/pca.pdf accessed October 7, 2013.

[Vek13] Olga Veksler. CS 434s/541a Pattern Recognition. http://www.csd.uwo.ca/~olga/Courses/CS434a_541a/Lecture7.pdf, accessed October 7, 2013.