CSC 480/580 Homework 3

Due: 4/15 (Mon) 5pm

Instructions:

- If you use math symbols, please define it clearly before you use it (unless they are standard from the lecture).
- You must provide the derivation for obtaining the answer and full source code for whatever problem you use programming. Please email your source codes to `csc580homeworks@gmail.com`.
- Please use the problem & subproblem numbering of this document; do not recreate or renumber them.
- Submit your homework on time to gradescope. NO LATE DAYS, NO LATE SUBMISSIONS ACCEPTED.
- The submission must be one single PDF file (use Acrobat Pro from the UA software library if you need to merge multiple PDFs).
- Please include your answers to all questions in your submission to Gradescope. (Do not store your answers in your source codes or Jupyter notebooks - I will not look at them by default.)
  - You can use word processing software like Microsoft Word or LaTeX.
  - You can also hand-write your answers and then scan it. If you use your phone camera, I recommend using TurboScan (smartphone app) or similar ones to avoid looking slanted or showing the background.
  - Watch the video and follow the instruction: `https://youtu.be/KMPoby5g_nE` .
- Collaboration policy: do not discuss answers with your classmates. You can discuss HW for the clarification or any math/programming issues at a high-level. If that is the case, please mention who you've talked to in your submission. Declaring your collaborators will not result in deduction of points; instead, failure to declare your collaborators counts as academic integrity violation.

Problem 1. Probabilistic Reasoning.
(a) Denote background evidence by event E. Suppose X, Y are two other events. Prove the conditional version of Bayes' rule:

$$P(X \mid Y, E) = \frac{P(Y \mid X, E)P(X \mid E)}{P(Y \mid E)}$$

(b) Consider the following Bayesian network:

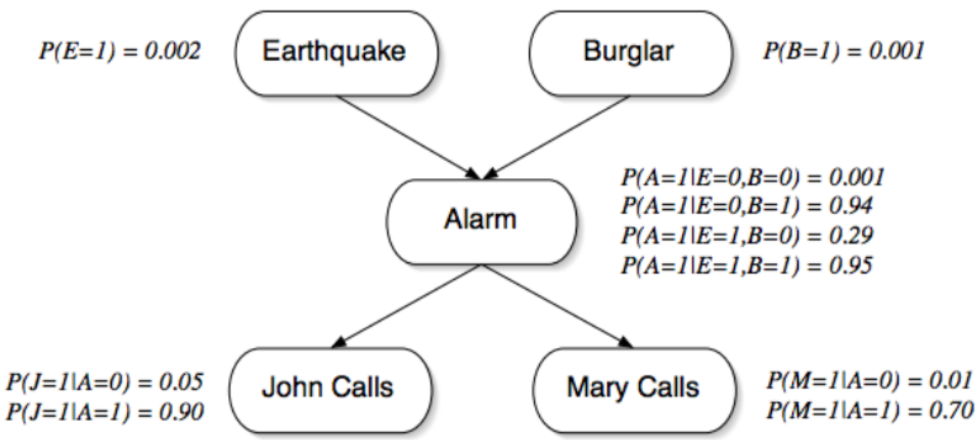

Figure 1: A Bayesian network for a house (picture credit: Prof. Lawrence Saul)

(i) Using Bayes' rule, calculate $P(E = 1 \mid A = 1)$. Is it larger than $P(E = 1)$? Does it make intuitive sense?
(ii) Using Bayes' rule, calculate $P(E = 1 \mid A = 1, B = 1)$. Is it larger than $P(E = 1 \mid A = 1)$? Someone comments that the event of burglary "explains away" the happening of an earthquake given the alarm sounding. Do you agree with this comment, why?
(iii) Calculate the joint distribution of (J, M). Is $J \perp\!\!\!\perp M$? Is $J \perp\!\!\!\perp M \mid A$? Justify your answers.

Problem 2. Principal Component Analysis

Download three.txt and eight.txt, which can be found in our Piazza page. Each has 200 handwritten digits. Each line is for a digit, vectorized from a 16x16 gray scale image.

(a) Each line has 256 numbers: they are pixel values (0=black, 255=white) vectorized from the image as the first column (top down), the second column, and so on. Visualize using python the two gray scale images corresponding to the first line in three.txt and the first line in eight.txt.

(b) Put the two data files together (threes first, eights next) to form a $n \times d$ matrix X where $n = 400$ digits and $d = 256$ pixels. The i-th row of X is $x_i^\top$, where $x_i \in \mathbb{R}^d$ is the i-th image in the combined data set. Compute the sample mean $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. Visualize $\bar{x}$ as a 16x16 gray scale image.

(c) Center X using $\bar{x}$ above. Then form the sample covariance matrix $S = \frac{X^\top X}{n-1}$. Show the 5x5 submatrix $S(1 \ldots 5, 1 \ldots 5)$.

(d) Use appropriate software/library to compute the two largest eigenvalues $\lambda_1 \geq \lambda_2$ and the corresponding eigenvectors v_1, v_2 of S. For example, in python one can use `scipy.sparse.linalg.eigs`. Show the value of λ_1, λ_2. Visualize v_1, v_2 as two 16x16 gray scale images. Hint: you may need to scale the values to be in the valid range of grayscale ([0, 255] or [0,1] depending on which function you use). You can shift and scale them in order to show a better picture. It is best if you can show an accompany 'colorbar' that maps gray scale to values.

(e) Now we project (the centered) X down to the two PCA directions. Let $V = [v_1, v_2]$ be the $d \times 2$ matrix. The projection is simply XV. (To be precise, these are the coefficients along the principal directions, not the projection itself.) Show the resulting two coordinates for the first line in three.txt and the first line in eight.txt, respectively.

(f) Report the average reconstruction error $\frac{1}{n}\sum_{i=1}^{n} \|x_i V V^\top - x_i\|^2$, where $x_i \in \mathbb{R}^{1 \times d}$ is the i-th row of the centered data matrix X.

(g) Now plot the 2D point cloud of the 400 digits after projection. For visual interest, color points in three.txt red and points in eight.txt blue. But keep in mind that PCA is an unsupervised learning method and it does not know such class labels.

Problem 3: Project check-in. Use this opportunity to continue making progress on your project. Give brief updates on the project for the following:

- What have you achieved?
- What difficulties have you encountered?
- What remains to be done? List all your todo items and give your deadline for each of them. (HW4 will be even shorter, to give you more time for your project.) Make sure you allocate 1-2 weeks for writing up the project report.