Stats 545: Midterm exam

This is a 75-minute exam for 32 points. Write your name and PUID on each sheet, and also include the number of answer sheets. Attempt all questions.

1 `R` code [6 pts]

Let `A` be an $N \times N$ `R` matrix, with elements `a_ij`, and `b` an N-dimensional vector.

1. What is the result of `A*b`? What is the result of `A %*% b`? [2 pts]
2. Write 1-2 lines of `R` to find the number of zeros in each column of `A`. Use vectorization for full points. [2 pts]
3. Consider the `R` statement `b[length(b)+1] <- 1`. What is its cost in big-O notation? Why? [2 pts]

2 Matrix operations [6 pts]

Let A, B, C be $N \times N$ invertible matrices.

1. What is the definition of the matrix norm $\|A\|_2$? If $A = BC$, show that $\|A\|_2 \leq \|B\|_2 \|C\|_2$ [3pts]
2. What is the QR decomposition of a matrix A. Explain briefly how that can be used to solve $Ax = b$. [3pts]

3 Dynamic programming [8 pts]

1. Write the cost of Kalman filtering in big-O notation, explaining what the terms are. [2pts]
2. Explain what the cost of sorting N elements is, using the heap sort. Justify this in 1-2 lines. [2pts]
3. Explain briefly why the k-means algorithm always converges. [2pts]
4. Explain briefly the advantage of EM over k-means for learning means of a mixture of Gaussians. [2pts]

4 Eigenvalues [4 pts]

1. Let A be a symmetric $N \times N$ matrix; recall an eigenvalue λ of A satisfies $Au = \lambda u$. Show that if λ is an eigenvalue of A, then λ^{-1} is an eigenvalue of A^{-1}. [2 pts]
2. Describe the steps of how would you use the power-method to find the *smallest* eigenvalue of A? [2 pts]

5 The EM algorithm [8 pts]

We observe data from a mixture of two 1-d Gaussians, both with mean 0, but with unknown variances σ_1^2 and σ_2^2. The first component has probability π, and the second $1 - \pi$. We observe N datapoints $X = (x_1, \ldots, x_N)$, let the latent-clusters be $C = (c_1, \ldots, c_N)$.

1. Write down the log joint-probability $\log p(X, C|\mu, \pi, \sigma_1^2, \sigma_2^2)$. [2pts]
2. Write down the EM lower-bound $\mathcal{F}(q, \boldsymbol{\sigma_1^2}, \boldsymbol{\sigma_2^2})$. [2pts]
3. For variances σ_1^2, σ_2^2, write the $q_i(c_i)$ that maximizes $\mathcal{F}$ for observation i. Given the q's how would you update σ_1^2? You don't have to derive these, but explain the intuition. [4pts]