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 [4 pts]

1. `X` is an $n \times d$ matrix, each row being a d-dimensional observation. `cls` is length-n vector of integers, giving cluster assignments for each observation. Write a few lines of `R` to calculate the means of each cluster. [2 pts]
2. A is an $n \times n$ matrix. Write a few lines of `R` to calculate A^{1050} in an efficient manner. [2 pts]

2 Matrix operations [9 pts]

Let A, B, C be $n \times m, m \times p$ and $p \times q$ matrices.

1. How many scalar multiplications are required to calculate $A \cdot B$? To calculate $A \cdot B \cdot C$? [3pts]

Let A be and $n \times n$ matrix.

2. Definite the matrix norm $\|A\|_2$? What is the condition number $\kappa(A)$? Show that $\kappa(A) \geq 1$. [3pts]
3. What is the Cholesky decomposition of A? Given its Cholesky decomposition, how would you simulate an n-dim Gaussian with mean m and covariance A? You only can generate standard normals using `rnorm`. [3pts]

3 Dynamic programming [6 pts]

1. In class, we tried implementing a priority queue with a linked-list, and saw that this has a worst case cost of $O(N)$. Briefly explain why this in so. What is the cost of implementing a priority queue with a heap? [2pts]
2. You work at a bank and have N kinds of notes, with values $(V_1, V_2, \cdots, V_N)$. You have unlimited notes of each type. Someone cashes a check of value S, and wants the money using as few notes as possible. Provide a dynamic program to solve this problem, providing a) How you relate any solution to simpler solutions, and b) How you initialize your dynamic program. What is the cost of your algorithm? [4pts]

4 Exponential family distributions [5 pts]

1. Give an example of a probability distribution that is *not* exponential family. Write down its density, and explain why it is not. [1 pts]
2. Write down Jensen's inequality, and explain the intuition behind it. Give an example where Jensen's inequality becomes an equality. [2 pts]
3. What is the Kullback-Liebler divergence between two densities p and q. Show that $\mathrm{KL}(p\|q) \geq 0$ (Hint: it might be easier to show $-\mathrm{KL}(p\|q) \leq 0$). [2 pts]

5 The EM algorithm [8 pts]

We observe data from a mixture of two 1-d Gaussians, both with variance 1, and with means μ and 2μ. 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 $\log p(X, C|\mu, \pi)$. [2pts]
2. Write down the EM lower-bound $\mathcal{F}(q, \pi, \mu)$. [2pts]
3. For μ and π given, write the $q_i(c_i)$ that maximizes $\mathcal{F}$ for observation i. Given the q's how would you update π and μ? You don't have to derive these, but explain the intuition. [4pts]