Stats 545: Midterm exam (75 minutes for 30 points)

Write your name and PUID on each sheet, and also include the number of answer sheets. Attempt all questions. The difficulty of a question need not be correlated with its score.

1 Miscellaneous topics [10 pts]

1. Define what a convex function is. Is maximum likelihood estimation for a fully observed exponential family model convex, concave, or problem-specific? Explain your answer very briefly. [2pts]
2. Define the entropy $H(q)$ of a distribution $q(x)$. Use Jensen's inequality to show $H(q) \geq 0$. [2pts]
3. Briefly explain why the cost of `heapsort` is $O(N \log N)$. Why is `quicksort` a more popular sorting algorithm, even though its cost is $O(N^2)$? [2pts]
4. Explain the motivation for k-medoids. What are top-down and bottom-up clustering? [2pts]
5. Briefly explain the statistical problem that the Kalman filter solves. [2pts]

2 Matrix operations [6 pts]

1. What is the Cholesky decomposition of a matrix A? Do all matrices have a Cholesky decomposition? Explain. Given its Cholesky decomposition, how will you calculate the determinant of A? [2pts]
2. Define the operator norm of a matrix A. Give its interpretation. Give a matrix (other than the identity or its negative) with norm equal to 1. You need not write out the elements of the matrix, just describe it. [2pts]
3. For a decomposition $A = BC$, show that $\|A\| \leq \|B\|\|C\|$. [2pts]

3 Dynamic programming [7 pts]

You have matrices $A_1, A_2, \ldots, A_N$. Matrix A_i has dimension $d_i \times d_{i+1}$ (so matrix A_1 is $d_1 \times d_2$, A_2 is $d_2 \times d_3$ etc).

1. What is the cost of multiplying $A_1 \cdot A_2$? What is the cost of multiplying $A_1 \cdot A_2 \cdot A_3$? [2pts]
2. Briefly explain why calculating **the cost** of multiplying $A_1 \cdot A_2 \cdot A_3 \ldots \cdot A_N$ is difficult for large N. [1pts]

We will use dynamic programming to solve the previous question. The forward pass will give the cost of the multiplication, and the backward pass will give how to do the multiplication. We only focus on the forward pass.

3. For integers $i < j < k$, suppose you know the cost of the products $A_i \cdot A_{i+1} \cdot \ldots \cdot A_j$ and $A_{j+1} \cdot \ldots \cdot A_k$ for all j between i and k. Then what is the cost of $A_i \cdot A_{i+1} \cdot \ldots \cdot A_k$? What is the cost of calculating this cost? [2pts]
4. Describe the overall dynamic programming algorithm to compute the cost of $A_1 \cdot A_2 \cdot A_3 \ldots \cdot A_N$. [2pts]

4 The EM algorithm [7 pts]

Someone has two coins with probability of heads equal to μ and π. They flip the μ-coin first. If heads `H`, they flip the π-coin twice, and report the outcomes. If tails `T`, they flip the π-coin just *once*, but report the outcome twice (i.e. if the π-coin gives heads, `H`, they report `HH`).

This experiment is repeated N times, but you only observe the outputs of the π-coin. Call your ith observation x_i, this equals `HH, TT, HT` or `TH`. Write c_i for the corresponding output of the μ-coin. Given a set of observations $X = \{x_1, \ldots, x_n\}$, you want to estimate the parameters μ and π.

1. A simple approach assumes that if x_i is `HH` or `TT`, then the μ-coin was tails ($c_i =$ `T`), else $c_i =$ `H`. If you fill in the missing c_i's this way, what are the estimates of π and μ? Why might this not be a good approach? [2pts]

Instead, we will use the EM algorithm.

2. Write down the log joint-probability $\log p(X, C|\mu, \pi)$. Write down the EM lower-bound $\mathcal{F}(q, \mu, \pi)$. Given μ and π, write the $q_i(c_i)$ that maximizes $\mathcal{F}$. Given the q_i's and π, write the update for μ. [5pts]