Stats 545: Midterm exam

This is a 2-hour exam for 60 points. Write your name and PUID on each sheet, and also include the number of answer sheets. Attempt all questions.

1 Matrix operations [8 pts]

Assume multiplying two $N \times N$ matrices requires N^3 operations while an $N \times N$ matrix times an $N \times 1$ vector requires N^2 operations. Let A be an $N \times N$ matrix, and b an $N \times 1$ vector. Define $A^4 = A \cdot A \cdot A \cdot A$.

1. How many operations are needed to calculate $A^4 \cdot b$? [2]
2. How many operations are needed to calculate A^4? [2]

Let L be an $N \times N$ lower triangular matrix (i.e. $L_{ij} = 0$ for $i < j$)

3. Write down an expression for its determinant $|L|$. [2]
4. What is the structure of L^{-1}? (e.g. lower/upper triangular, diagonal, unstructured etc.) [2]

2 Binary heap [14 pts]

Consider a binary heap with N nodes, the root containing the maximum (see the left subfigure for an example). In big-O notation (you can ignore multiplicative constants), give

1. the cost of **finding** the **largest** element in a binary heap. [2]
2. the cost of **removing** the **largest** element from a binary heap. [2]
3. the cost of **finding** the **fifth largest** element in a binary heap ($N > 5$). [2]
4. the cost of **finding** the **smallest** element in a binary heap. [2]

We will implement the heap as a simple list, organizing its elements sequentially as in the right subfigure.

5. What is the index of the parent node of element i of the list? [1]
6. What are the indices of the left and right child nodes of element i of the list? [2]
7. Write a few lines of pseudocode showing how you would insert a new element `new_elem` into the heap as implemented above. You are given the list `my_heap` and its length `heap_len`. Thus, `my_heap[i]` indexes the ith element of the heap. [3]

3 Hidden Markov models [14 pts]

Consider the same setup as the homework: an N-state Markov chain whose state at time t is S_t. The initial distribution over states is π^1, and the transition matrix is A with $P(S_{t+1} = j|S_t = i) = A_{ij}$. At time t we observe Y_t from the distribution $P_Y(Y_t|S_t)$. We have T observations $(Y_1, \ldots, Y_T)$.

In the homework we were interested in calculating marginals $P(S_t|Y_1, \ldots, Y_T)$. Now we wish to find the most likely sequence of states $(S_1^*, \ldots, S_T^*) = \text{argmax } P(S_1, \ldots, S_T|Y_1, \ldots, Y_T)$.

1. What is the cost of brute force search in big-O notation as a function of $N$and T? [1]

Instead, we use dynamic programming. We modify the Baum-Welch forward pass, defining α-messages as:

$$\alpha_t(i) = \max_{(S_1,\ldots,S_{t-1})} P(S_1,\ldots,S_{t-1},S_t = i, Y_1,\ldots,Y_t)$$

Thus, $\alpha_t(i)$ looks at the joint probabilities (with $Y_1,\ldots,Y_t$) of all sequences ending with $S_t = i$, and returns the largest probability.

2. Write down $\alpha_1(i)$ as a function of π^1 and $P_Y(Y_1|S_1)$. [2]

3. Write down $\alpha_{t+1}(i)$ as a function of α_t, A and P_Y. [5]

4. At the end of the forward pass, we have $\alpha_T(i),\ i \in \{1,\ldots,N\}$. What will you set S_T to? [2]

5. The backward pass successively sets S_t given S_{t+1}. Write down how you will set S_t. [4]

4 The EM algorithm [14 pts]

We observe data from a mixture of two 1-d Gaussians with unknown means μ_1 and μ_2, and common variance σ^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 $\log p(X, C|\mu,\pi,\sigma^2)$. [2]

2. Write down the EM lower-bound $\mathcal{F}(q,\boldsymbol{\mu})$. [3]

3. For means μ_1,μ_2, write the $q_i(c_i)$ that maximizes $\mathcal{F}$ for observation i. Explain how you got this. [3]

4. Given the q_i's how would you update μ_1? You don't have to derive this, but explain the intuition. [3]

5. What happens to the q_i of part 3 as $\sigma^2 \to 0$. Explain qualitatively. Also explain qualitatively what happens to the overall EM algorithm. [3]

5 Monte Carlo sampling [10 pts]

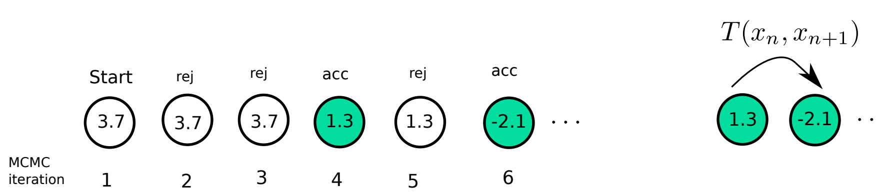
Your friend wants samples from a distribution $\pi(x)$. He uses a Metropolis-Hastings sampler with proposal distribution $q(x^*|x_n) = q(x^*)$. Observe that the new proposal is *independent* of the current state x_n.

1. Write the acceptance probability $\alpha(x_n, x^*)$ of a move from x_n to x^* in terms of $\pi(x)$ and $q(x)$. [2]

2. For this special case where the proposals are independent of the current state, are the resulting MCMC samples $(x_1, x_2, \ldots)$ also independent variables, or do they still form a Markov chain? Explain why. [2]

3. Assume that π and q small, so that directly calculating the acceptance probability will lead to numerical issues like `NA`s. Much better is to use $\log\pi$ and $\log q$. Write one or two lines of pseudocode for how you will accept or reject a move from x_n to x^*. You can assume you have access to a random number generator `randu`, which samples from Unif$(0,1)$. [3]

4.



Your friend is confused about Metropolis-Hastings. Instead of keeping *all* samples in the chain, he only keeps the accepted proposals (filled circles in the figure). What is the transition probability $T(x_n, x_{n+1}) = P(x_{n+1}|x_n)$ from one accepted sample to the next? [3]