# 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 Matrix operations [8 pts]

Let $A, Q, R$ be $N \times N$ matrices and $x$ an $N \times 1$ vector.

1. What is the definition of the matrix norm $\|A\|_2$? If $A$ has an eigenvalue $\lambda$, show that $\|A\|_2 \geq \lambda$. If $A = QR$, where $Q$ is an orthonormal (rotation) matrix, show that $\|Q\|_2 = 1$ and $\|A\|_2 = \|R\|_2$. [3pts]
2. If matrix-matrix and matrix-vector multiplication take $N^3$ and $N^2$ operations, how many operations are needed to calculate $A^8 x$ and $A^9$? What is the cost of $A^P$ in big-O, as a function of $N$ and $P$? [3pts]
3. Define QR and LU decomposition, and give an advantage of each over the other. [2pts]

## 2 Sorting and gradient descent [8 pts]

1. Describe an algorithm to find the median (i.e. middle value) of $N$ numbers, specifying its cost in big-O notation. Your algorithm should be faster than $O(N^2)$. You can assume $N$ is odd and all numbers are unique. [3pts]
2. Give the big-O cost of quicksort. Give its average cost, explaining what you mean by 'average'. [2pts]
3. Give an advantage and disadvantage of gradient descent vs stochastic gradient descent. [1pt]
4. Briefly explain the backward Euler (or proximal point) method. Explain why it never increases the value of the objective function. [2pts]

## 3 Clustering and dynamic programming [8 pts]

1. Explain single linkage and centroid linkage. For two clusters of size $M$ and $N$, what is the cost of computing single linkage distance? And centroid linkage? [3pts]

There are $N$ houses in a row, the $i$th house having value $v_i$ in it. A thief wants to steal from these houses, but cannot steal from two adjacent houses because the owner of a stolen house will wake up their left and right neighbors. The thief wants to steal as much value as possible.

2. A greedy strategy is to steal from the most valuable house, and then from the next most valuable house that is available, repeating till no houses are left. Give a setting of $N$ and $v_i$'s where this is suboptimal. [1pts]
3. Instead, the thief will use dynamic programming to decide which houses to break into. Write $S_i$ for the maximum value he can steal from the first $i$ houses. Express $S_i$ as a function of $S_j, j < i$. Use this to describe how to compute $S_N$ (the forward pass), and then to decide which houses to break into (the backward pass). Write the cost of this algorithm in big-O notation, explaining your answer. [4pts]

## 4 The EM algorithm [8 pts]

1. What is a *minimal* exponential family distribution. Give an example that is NOT minimal. [1pts]
2. Give an advantage and disadvantage of EM vs gradient ascent to perform MLE. [1pts]
3. Define KL divergence between probabilities $p(x)$ and $q(x)$. Prove $\mathrm{KL}(p\|q) \geq 0$ using Jensen's inequality.[2pts]

$x_i$ follows a Poisson distribution with with unknown mean $\lambda$ (recall that the Poisson distribution has the form $p(x|\lambda) = \lambda^x \exp(-\lambda)/x!$). Instead of $x_i$, we observe $y_i$, where $y_i = x_i$ if $x_i \geq 5$, else $y_i = 0$. We observe $N$ datapoints $Y = (y_1, \ldots, y_N)$.

1. Write the log joint-probability $\log p(Y, X|\lambda)$. Given $\lambda, \sigma^2$, write the $q_i(x_i)$ that maximizes $\mathcal{F}$. Is this an exponential family distribution? Explain. Given the $q$'s how do you update $\lambda$? [4pts]