CSE 151: Introduction to Machine Learning	Spring 2018
Problem Set 1	
Instructor: Kamalika Chaudhuri	**Due on:** never

Problem 1 (10 points)

Let u_1 and u_2 be vectors such that $\|u_1\| = \|u_2\| = 1$, and $\langle u_1, u_2 \rangle = 0$. For any vector x, we define $P(x)$ as the vector $P(x) = \langle x, u_1 \rangle u_1 + \langle x, u_2 \rangle u_2$.

1. How would you geometrically interpret $P(x)$? (Hint: Think about projections)
2. Show that: $\|P(x)\|^2 = \langle x, u_1 \rangle^2 + \langle x, u_2 \rangle^2$.
3. Using parts (1) and (2), show that $\|P(x)\| \leq \|x\|$. When is $\|P(x)\| = \|x\|$?

Problem 2 (10 points)

Given two column vectors x and y in d-dimensional space, the outer product of x and y is defined to be the $d \times d$ matrix $x \circ y = xy^\top$.

1. Show that for any x and y, $x^\top (x \circ y) y = \|x\|^2 \|y\|^2$. When is this equal to $x^\top \langle x, y \rangle y$?
2. Show that for any non-zero x and y, the outer product $x \circ y$ always has rank 1.
3. Let $x_1, \ldots, x_n$ be n $d \times 1$ data vectors, and let X be the $n \times d$ data matrix whose i-th row is the row vector $x_i^\top$. Show that:

$$X^\top X = \sum_{i=1}^{n} x_i \circ x_i$$

Problem 3 (10 points)

Suppose A and B are $d \times d$ matrices which are symmetric (in the sense that $A_{ij} = A_{ji}$ and $B_{ij} = B_{ji}$ for all i and j) and positive semi-definite. Also suppose that u is a $d \times 1$ vector such that $\|u\| = 1$. Which of the following matrices are always positive semi-definite, no matter what A, B and u are? Justify your answer.

1. $10A$.
2. $A + B$.
3. $uu^\top$.
4. $A - B$.
5. $I - uu^\top$ (Hint: Write down $x^\top (I - uu^\top) x$ in terms of some dot-products, and try usng Cauchy-Schwartz.)

Problem 4 (10 points)

In class, we discussed how to define a *norm* or a *length* for a vector. It turns out that one can also define a norm or a length for a matrix. Two popular matrix norms are the Frobenius norm and the spectral norm. The Frobenius norm of a $m \times n$ matrix A, denoted by $\|A\|_F$ is defined as:

$$\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2}$$

The spectral norm of a $m \times n$ matrix A, denoted by $\|A\|$ is defined as:

$$\|A\| = \max_{x \neq \mathbf{0}} \frac{\|Ax\|}{\|x\|}$$

where x is a $n \times 1$ vector.

1. Let I be the $n \times n$ identity matrix. What is its Frobenius norm? What is its spectral norm? Justify your answer.
2. Suppose $A = uv^\top$ where u is a $m \times 1$ vector and v is a $n \times 1$ vector. Write down the Frobenius norm of A as function of $\|u\|$ and $\|v\|$. Justify your answer.
3. Write down the spectral norm of A in terms of $\|u\|$ and $\|v\|$. Justify your answer.

Problem 5 (10 points)

Let x be a $d \times 1$ vector. Let y_i be constants, z_i be $d \times 1$ constant vectors, and β_i be $d \times 1$ constant vectors for $1 \leq i \leq n$. Write down the gradients for each of the following multivariate functions with respect to x. Given the other parameters describing the function, what is the time required to compute the gradient at a specific value of x?

1. $F(x) = \sum_{i=1}^{n} \log(1 + e^{-y_i x^\top z_i})$.
2. $G(x) = \sum_{i=1}^{n} (x^\top \beta_i - y_i)^2$.
3. $H(x) = \sum_{i=1}^{n} x_i \log \frac{1}{x_i}$.
4. $J(x) = \log(\sum_{i=1}^{n} e^{2x_i})$.