Homework 0, due Monday September 12 — COMS 4772 Fall 2016

Problem 1 (Values)**.** Name one or two of your own personal, academic, or career values, and explain how you hope CS/ML theory can be of service to those values.

Problem 2 (Stuff you must know). The course website `http://www.cs.columbia.edu/~djhsu/coms4772-f16/` has information about the course prerequisites, course requirements, academic rules of conduct, and other information. You are required to understand this information and abide by the rules of conduct, regardless of whether or not you can solve the following problems.

(a) True or false: I may share my homework write-up with another student as long as (1) the write-up only contains solutions for at most half of the problems, (2) we list each other as discussion partners on the submitted write-up.

(b) True or false: I may use any outside reference material to help me solve the homework problems as long as I appropriately acknowledge these materials in the submitted write-up.

Problem 3 (Stuff you should know)**.** Let $\boldsymbol{A}$ be the 6×6 matrix defined as follows (by a product of three 6×6 matrices):

$$\boldsymbol{A} := \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \frac{1}{4} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 \\ 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}.$$

(a) Let R be the span of $\boldsymbol{A}$'s rows. What is the dimension of R?

(b) $\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \\ 2 \end{bmatrix}$ an eigenvector of $\boldsymbol{A}$. True or false?

(c) Every eigenvector of $\boldsymbol{A}$ corresponding to eigenvalue $1/2$ has the form $\begin{bmatrix} 0 \\ c \\ c \\ 0 \\ 0 \\ 0 \end{bmatrix}$ or $\begin{bmatrix} 0 \\ c \\ -c \\ 0 \\ 0 \\ 0 \end{bmatrix}$ for some real number c. True or false?

(d) Let V be the subspace spanned by eigenvectors of $\boldsymbol{A}$ corresponding to the eigenvalue $1/4$. What is the dimension of V?

(e) What is the largest eigenvalue of $\boldsymbol{A}^3$?

Problem 4 (Random stuff you should know)**.**

(a) A biased coin with $P(\text{heads}) = 1/5$ is tossed repeatedly until heads comes up. What is the expected number of tosses?

(b) You create a random sentence of length n by repeatedly picking words at random from the vocabulary $\{\text{a}, \text{is}, \text{not}, \text{rose}\}$, with each word being equally likely to be picked. What is the expected number of times that the phrase "a rose is a rose" will appear in the sentence? (*Note*: the appearances may overlap.)

(c) A permutation π is picked uniformly at random from the space of all permutations on a non-empty set T. Let X be the number of *fixed points* of π—i.e., the number of $t \in T$ such that $\pi(t) = t$. What is the expected value of X^2?

(d) Let X be a random variable with density p given by

$$p(x) \ := \ \begin{cases} 0 & \text{if } x < 0 \,, \\ \lambda e^{-\lambda x} & \text{if } x \geq 0 \,. \end{cases}$$

Here, λ is a positive number (typically called the rate parameter). If $P(X \leq 1000000) = 0.5$, then what is the value of λ?

(e) Suppose the pair of random variables (X_1, X_2) has probability density function p given by

$$p(x_1, x_2) \ := \ \begin{cases} c & \text{if } 0 \leq x_1 \leq 0.5 \text{ and } 0 \leq x_2 \leq 1 \,, \\ 0 & \text{otherwise} \,. \end{cases}$$

Here, c is a constant (that does not depend on x_1 or x_2). What should be the value of c so that p is a valid probability density function?

(f) Continuing from (e), what is the probability that $X_2 \geq X_1$?

(g) Continuing from (e), define another random variable Y on the same probability space as X_1 and X_2 by

$$Y \ := \ \begin{cases} 1 & \text{if } X_1 > 2X_2 \,, \\ -1 & \text{otherwise} \,. \end{cases}$$

Are X_1 and Y independent? What is the expected value of Y?

(h) Continuing from (e), define yet another random variable Z on the same probability space as X_1 and X_2 by

$$Z \ := \ \begin{cases} 1 & \text{if } X_2 > 1/2 \,, \\ -1 & \text{otherwise} \,. \end{cases}$$

Are X_1 and Z independent? What is the expected value of $X_1 Z$?

Problem 5 (Stuff you should be able to prove)**.** The *convex hull* of a set $S \subseteq \mathbb{R}^d$, denoted $\operatorname{conv}(S)$, is defined by

$$\operatorname{conv}(S) \;:=\; \left\{ \sum_{i=1}^{k} p_i \boldsymbol{x}_i : k \in \mathbb{N},\, \boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_k \in S,\, p_1, p_2, \dots, p_k \in (0,1]\,,\, \sum_{i=1}^{k} p_i = 1 \right\}.$$

In other words, $\operatorname{conv}(S)$ is the set of all *convex combinations* of points in S.

In this problem, you will prove that for any set $S \subseteq \mathbb{R}^d$, any $\varepsilon \in (0,1)$, and any point $\boldsymbol{z}$ in the convex hull of a set $S \subseteq \mathbb{R}^d$, there is a convex combination $\boldsymbol{y} \in \operatorname{conv}(S)$ of at most $\lceil 1/\varepsilon^2 \rceil$ points in S such that $\|\boldsymbol{y} - \boldsymbol{z}\|_2 \leq \varepsilon\Delta$, where $\Delta := \max_{\boldsymbol{x},\boldsymbol{y} \in S} \|\boldsymbol{x} - \boldsymbol{y}\|_2$.

The following procedure takes as input the set S and a point $\boldsymbol{z} \in \operatorname{conv}(S)$, and constructs the required point $\boldsymbol{y} \in \operatorname{conv}(S)$. Let $T := \lceil 1/\varepsilon^2 \rceil$. Start with $\boldsymbol{x}_1$ being any point in S. Then, for $t = 2, 3, \dots, T$, let $\boldsymbol{x}_t$ be any point in S such that $\left\langle \sum_{i=1}^{t-1} (\boldsymbol{x}_i - \boldsymbol{z}), \boldsymbol{x}_t - \boldsymbol{z} \right\rangle \leq 0$ (the existence of a suitable point $\boldsymbol{x}_t$ requires proof). Finally, return $\boldsymbol{y} := \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{x}_t$.

(a) Prove the following claim. Pick any $S \subseteq \mathbb{R}^d$ with a point $\boldsymbol{q} \in \operatorname{conv}(S)$. For any $\boldsymbol{r} \in \mathbb{R}^d$, there exists $\boldsymbol{x} \in S$ such that $\langle \boldsymbol{r} - \boldsymbol{q}, \boldsymbol{x} - \boldsymbol{q} \rangle \leq 0$. (This claims explains why it is possible to pick a point $\boldsymbol{x}_t$ in step t of the algorithm.)

(b) Prove that $\|\boldsymbol{z} - \boldsymbol{y}\|_2^2 \leq \Delta^2 / T$.