Homework 01: Linear Algebra, Probability, and Statistics

2021 Spring, CS 6501 Natural Language Processing

Due on **Feb. 23, 11:59 PM**

1. **Matrix-Vector Multiplication** (3 points) Solve the following matrix-vector multiplication problem

$$\boldsymbol{M} \cdot \boldsymbol{x} \tag{1}$$

where

$$\boldsymbol{M} = \begin{bmatrix} 0.1 & 0.3 & 0.7 & 0.4 & 0.2 \\ 0.8 & 1.0 & 0.6 & 0.1 & 0.9 \end{bmatrix} \tag{2}$$

and

$$\boldsymbol{x}^{\mathsf{T}} = [1, 0, 2, 1, 1] \tag{3}$$

   - (1 point) Solve this problem by hand. Please show the procedure of your calculation.
   - (2 points) Verify your result with a simple code write in PyTorch/Tensorflow. To do this, you need to be familiar with the PyTorch/Tensorflow functions of creating tensors from arrays. Please submit your code with name [`CompingID-Q1`].py or [`CompingID-Q1`].ipynb if you use IPython Notebook.

2. **Special Matrices** (3 points) PyTorch/Tensorflow provides functions to create special matrices, including orthogonal matrices, symmetric metrices, identity metrices.

   - (1 point) Create an 4×4 orthogonal matrix $\boldsymbol{U}$ using PyTorch/Tensorflow functions
   - (2 points) Write a code to verify that $\boldsymbol{U} \cdot \boldsymbol{U}^{\mathsf{T}} = \boldsymbol{U}^{\mathsf{T}} \cdot \boldsymbol{U} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix

   Please submit your code with name [`CompingID-Q2`].py or [`CompingID-Q2`].ipynb if you use IPython Notebook.

3. **Geometric Meaning of Matrix-Vector Multiplication** (2 points) For any matrix $\boldsymbol{A} \in \mathbb{R}^{n \times n}$ and vector $\boldsymbol{x} \in \mathbb{R}^n$, matrix-vector multiplication $\boldsymbol{Ax}$ can be considered as a linear transformation of $\boldsymbol{x}$.

   Please prove that, if $\boldsymbol{A}$ is an orthogonal matrix, this linear transformation geometrically is a rotation around the origin. In other words, the proof needs to show that $\boldsymbol{Ax}$ does not change the "length" of $\boldsymbol{x}$, aka, $\|\boldsymbol{Ax}\|_2 = \|\boldsymbol{x}\|_2$. [Hint: you may need the following property of matrix (vector) transpose: $(\boldsymbol{Ax})^{\mathsf{T}} = \boldsymbol{x}^{\mathsf{T}}\boldsymbol{A}^{\mathsf{T}}$]

4. **Random Event and its Sample Spaces** (4 points) In our lecture, we talked about several examples of random events, such as *flipping a coin* and *predicting the weather tomorrow*. In this problem, please pick a real-world *NLP* example as the random event, then

   - (1 point) define a random variable related to this random event
   - (1 points) define the sample space of this random variable and justify the size of the sample space
   - (2 point) define a probability distribution over the sample space. To answer this question, you need to specify (1) what this distribution is? (2) what the parameters it has? and (3) what each parameter means in the defined distribution? You don't have to specify the values of these parameters.

5. **Dependence of Two Random Variables** (2 points) Given two random variable X, Y and the joint probability distribution,

	$X=0$	$X=1$
$Y=0$	0.12	0.28
$Y=1$	0.18	0.42

Are these two random variable independent? Please justify your answer.

6. **MLE on Bernoulli Distributions** (2 points) On page 44 of the lecture slides, we give the definition of the log-likelihood function of a Bernoulli distribution with the obversations $\{x^{(1)}, \ldots, x^{(n)}\}$,

$$\ell(\theta) = \sum_{i=1}^{n} \left\{ x^{(i)} \log \theta + (1 - x^{(i)}) \log(1-\theta) \right\} \tag{4}$$

The way to find the value of parameter θ is to maximize the (log-)likelihood function. In this case, it involves solving the following problem

$$\frac{d\ell(\theta)}{d\theta} = 0. \tag{5}$$

Please show the answer to equation 5 is

$$\theta = \frac{1}{n} \sum_{i=1}^{n} x^{(i)} \tag{6}$$