CSE 404: Introduction to Machine Learning— Self Assessment

Instructor: Jiayu Zhou

Machine learning is an interdisciplinary field that involves the following topics: linear algebra, probability, and numerical optimization. In this class you will also be required to use PYTHON for programming. Before taking the class, please use the following questions to check if you have enough background knowledge.

1. Given two column vectors $a = \begin{bmatrix} 1 \\ 2 \\ 4 \end{bmatrix}$ and $b = \begin{bmatrix} 3 \\ -5 \\ 1 \end{bmatrix}$, what is the value of the inner product of the two vectors? Compute by hand and show the intermediate steps.

2. What is matrix multiplication? Can you multiply two arbitrary matrices? Given two matrices $A = \begin{bmatrix} 1 & 4 & -3 \\ 2 & -1 & 3 \end{bmatrix}$ and $B = \begin{bmatrix} -2 & 0 & 5 \\ 0 & -1 & 4 \end{bmatrix}$, can you multiply the two matrices? How about A^T and B? When multiplication is possible, compute by hand and show the intermediate steps. What is the *rank* of the resulting matrix?

3. In a study of Boy Scouts and Juvenile Delinquency[1], researchers have obtained the following statistics from past years:

Boy Scout	Delinquent	
	Yes	No
Yes	33	343
No	64	360

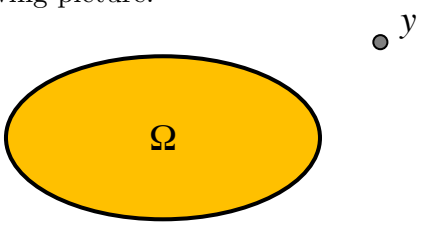
   (a) What is the probability of juvenile delinquent?
   (b) Given a boy scout, what is the probability of juvenile delinquent?
   (c) Is the status of boy scout independent of juvenile delinquent?

4. An *Euclidean projection* of a d-dimensional point $y \in \mathbb{R}^d$ to a set Ω is given by the following optimization problem:

$$x^* = \arg\min_x \|x - y\|_2^2, \quad \text{subject to: } x \in \Omega \tag{1}$$

   where Ω is the *feasible set*, $\|\cdot\|_2$ is the ℓ_2 norm of a vector, and $x^* \in \mathbb{R}^d$ is the projected vector.

   (a) What is x^* if $y = 1.1$ and $\Omega = \mathbb{N}$, where $\mathbb{N}$ is the set of natural numbers.
   (b) Locate the x^* in the following picture:



5. Use PYTHON with NUMPY/SCIPY package to verify your solution in Questions 1 and 2.

[1] https://onlinecourses.science.psu.edu/stat504/node/103