Data Science Homework 1

20221005

Submission Policy

- Deadline: 10/19, 23:59
  - Late submission will get no points
- Submit your file to the homework section of the NTU COOL system.
- File format: .pdf
  - Please make sure your file can be opened. A broken file will get no points.

Problem 1 Basic Definitions

- Let μ be a function from set Σ to R. That is, $\mu: \Sigma \rightarrow R$. What are the conditions for μ to be called a measure on Σ? (10%)
  - The one pre-condition that Σ must satisfy.
  - The 3 conditions that μ must satisfy.
- Let f() be a function from $X \times X$ to non-negative real numbers. What are the four conditions that f must satisfy for f() to be considered a metric of X? (10%)

Problem 2 Random Variable Transformation

Min() and Max() appears frequently in applications of data science.

Let X and Y be two independent random variables with identical probability density function given by

$$f(x) = \begin{cases} e^{-x} & for\ x > 0 \\ 0 & elsewhere. \end{cases}$$

(1) What is the probability density function of Z=max(X,Y) ? (10%)
(2) What is the probability density function of W=min(X,Y) ? (10%)

(Hint: “max(X, Y) ≤ a” means “X ≤ a and Y ≤ a”.)

Problem 3 Random Variable Transformatin

- Let X be a random variable whose distribution is Gamma($\boldsymbol{\theta}$=1, $\boldsymbol{\alpha}$=1).
    - Drive the distribution of 2X? (5%)
    - Expressive the distribution of 2X in terms of (1) Gamma distribution (2) Chi^2 distribution (that is, it is of Gamma distribution with what parameters and Chi^2 distribution with what parameter) (5%)

Problem 4 Statistical Distance

- Let x, y be two points (two vectors) in space. The function $d(x, y) = (x-y)^2$ is called the squared Euclidean distance. Show that the squared Euclidean distance is a Bregman divergence (10%) (hint: what is the F() in this Bregman divergence?)
- You are given an artificial neural network. The network implements a function that takes an input x and produce an output y. That is, it implements $y = F(x)$. Prove that the entropy of the output of this neural network will always be equal or less that of the input. (10%)

Problem 5 Point Estimation

- We have a population X whose distribution is uniform over the interval $(0,\theta)$. The prior distribution of θ is uniform over the interval $(0,1)$. Please derive the estimator of θ based on a sample of size $n \geq 2$, using:
  - The moment method (5%)
  - The MLE method (5%)
  - The MAP method (10%)
  - The Bayesian method using squared error loss function (10%)