Data Science Homework 2

20221020

Submission Policy

- Deadline: 11/3, 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 Goodness of Estimation

- Let
  - θ_1 and θ_2 be two unbiased estimators of θ, with $Var(\theta_1)=1$, $Var(\theta_2)=2$ and $Cov(\theta_1, \theta_2)=1/4$.
  - θ_3 and θ_4 be two unbiased estimators of θ, with $Var(\theta_3)=1$, $Var(\theta_4)=2$ and $Cov(\theta_3, \theta_4)=3/4$.
- What is the unbiased estimator with the lowest variance that you can construct from a linear combination of θ_1 and θ_2, and what's its variance? (10%)
- Answer the same questions for θ_3 and θ_4 (10%)
- Note:
  - You can observe which combination produces a new estimator with lower variance (for your own amusement, no additional points awarded here)

Problem 2 Interval Estimation

- Let X1, X2, ..., Xn be a random sample from a population with density function

  $$f(x;\theta) = e^{-(x-\theta)},\ if\, \theta < x < \infty$$

  where θ∈R is an unknown parameter. Show that Q = X(1) - θ is a pivotal quantity (10%). Use this pivotal quantity find a 100(1-α)% confidence interval for θ. (10%)

Problem 3 Type I and Type II Errors

- Suppose X has the density function

  $$f(x;\theta) = \begin{cases} \frac{1}{\theta} & \text{for } 0 < x < \theta \\ 0 & \text{otherwise} \end{cases}$$

  The null hypothesis is H_0: $\theta = 1$ against the alternative hypothesis H_a: $\theta = 2$. Suppose one observation of X is taken, what are the probabilities of Type I error (10%) and Type II error (10%) in testing if H_0 is rejected for $X > 0.92$.

Problem 4 Hypothesis Testing

- Let $X_1, X_2, ..., X_8$ be a random sample of size 8 from a Poisson distribution with parameter λ. The null hypothesis is H_0: $\lambda = 0.5$. Reject H_0 if the observed sum

$$\sum_{i=1}^{8} x_i \geq 8$$

  (1) Compute the significance level α of the test. (10%) (2) Write out the formula of the power function $\beta(\lambda)$ of this test. (10%)
- Note:
  - You may need to look up Poisson distribution and its meaning first, if you are not already familiar with it.

Poisson Distribution Table

λ =	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0	4.5	5.0
X=0	0.6065	0.3679	0.2231	0.1353	0.0821	0.0498	0.0302	0.0183	0.0111	0.0067
1	0.9098	0.7358	0.5578	0.4060	0.2873	0.1991	0.1359	0.0916	0.0611	0.0404
2	0.9856	0.9197	0.9197	0.8088	0.6767	0.5438	0.4232	0.3208	0.2381	0.1247
3	0.9982	0.9810	0.9344	0.8571	0.7576	0.6472	0.5366	0.4335	0.3423	0.2650
4	0.9998	0.9963	0.9814	0.9473	0.8912	0.8153	0.7254	0.6288	0.5321	0.4405
5	1.0000	0.9994	0.9994	0.9955	0.9834	0.9161	0.8576	0.7851	0.7029	0.6160
6	1.0000	0.9999	0.9991	0.9955	0.9858	0.9665	0.9347	0.8893	0.8311	0.7622
7	1.0000	1.0000	0.9998	0.9989	0.9958	0.9881	0.9733	0.9489	0.9134	0.8666
8	1.0000	1.0000	1.0000	0.9998	0.9989	0.9962	0.9901	0.9786	0.9597	0.9319
9	1.0000	1.0000	1.0000	1.0000	0.9997	0.9989	0.9967	0.9919	0.9829	0.9682
10	1.0000	1.0000	1.0000	1.0000	0.9999	0.9997	0.9990	0.9972	0.9933	0.9863

Problem 5 LDA

- We are given a sample of data belonging to two classes, in which each data point has two attributes.
    - The data points belonging to class 1 are: {(5, 3), (3, 5), (3, 4), (4, 5), (4, 7), (5, 6)}
    - The data points belonging to class 2 are: {(9, 10), (7, 7), (8, 5), (8, 8), (7, 2), (10, 8)}
- Perform LDA and find out the optimal projection vector (normalized to unit length), and its corresponding eigenvalue. Show your calculation steps. (20%)
- Note:
    - You are encouraged to write a program to calculate the answer. If you write a program, you are allowed to call function calls to do matrix operations, such as matrix inversion, etc., but you are not allowed not call any function call that calculates LDA in one step.
    - If you are not familiar enough with programming at this point, you are also allowed to use hand calculation, or use calculator to derive the answer.