Data Science 110 Homework 2

2021.10.22

Submission

- Deadline: 11/3 Wed. 23:59
- Submission delay: will get no points
- Upload: Ceiba homework section
- File format: PDF
- Format error: -10%

Problem 1 Type I and II errors (20%)

- Five trials $X_1, X_2, \ldots, X_5$ of a Bernoulli experiment were conducted to test H_o: p $= \frac{1}{2}$ against H_α: p $= \frac{3}{4}$. The null hypothesis H_o will be rejected if $\Sigma_{i=1}^{5} X_i = 5$.

Find out the probability of

(1)(10%) Type I error

(2)(10%) Type II error.

Problem 2 Hypothesis testing (20%)

- A manufacturer of car batteries claims that the life of his batteries is normally distributed with a standard deviation equal to 0.9 year.

- You took a random sample of size 10 of these batteries to examine his claim, and you found out this sample has a standard deviation of 1.2 years.

- If you decide to use a level of significance of 0.05, should you accept the manufacturer's claim that the standard deviation = 0.9 year, or should you believe the standard deviation > 0.9? Show you derivation and your answer.

- Hint: this is a one-tailed test.

Problem 3 Markov Properties (20%)

- Write down the following definition clearly:
  - What are the pairwise, local and global Markov properties for Markov random fields? (4x3%)
  - What is the local Markov properties for Bayesian network? (4%)
  - What is a Markov blanket? (4%)

Problem 4 LDA (40%)

- We are given a sample of data belonging to two classes. Each data point in the sample has two attributes.
  - The sample points belonging to class 1 are: {(5, 3), (3, 5), (3, 4), (4, 5), (4, 7), (5, 6)}
  - The sample 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 vectors (normalized to unit length), and its corresponding eigenvalues. Show your calculation steps.
- 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.

- Please calculate the sample variance with $\frac{\Sigma_{i=1}^{n} (x_i - \bar{x})^2}{n-1}$ (rather than n in the denominator)