

Home Assignment 3.

1. Problem

Vasiliy loves plov. During last 140 days he had eaten 30 plovs with lamb, 30 plovs with beef and 80 plovs with chicken.

Test the null hypothesis that all types of plov have the same probability using the likelihood ratio test.

2. Problem

Suppose we have data on daily web-site visits for registered and non - registered users:

Number of visits	Registered	Non-registered
Less than 5	80	87
More or equal to 5	53	86

Check if the association between two variables exist using Chi-squared test.

Measure of the strength of association using phi coefficient.

3. Problem

Consider linear discriminant analysis model with two classes and two explanatory variables $x = (a, b)$.

The model is given by

$$(x \mid y = 1) \sim \mathcal{N}(\mu_1; \Sigma), \quad (x \mid y = 2) \sim \mathcal{N}(\mu_2; \Sigma), \quad \mathbb{P}(y = 1) = 0.3,$$

where

$$\mu_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \mu_2 = \begin{pmatrix} 3 \\ 8 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 4 & -2 \\ -2 & 19 \end{pmatrix}.$$

Find the line $k_a \cdot a + k_b \cdot b = 1$ that separates the two classes, i.e. any point on the line has equal conditional probabilities to be in each class.

4. Problem

We want to conduct LDA analysis to determine number of visits to “Hilton Hotel” during the season (1, 2 or 3 times) based on 5 explanatory variables.

There are 200 observations.

The eigenvalues for discriminant functions are $\gamma_1 = 0.104$ and $\gamma_2 = 0.1395$. Test whether the groups are significantly different from each other.

Calculate the p-value for the corresponding test.

5. Problem

Maria has estimated logit model,

$$\mathbb{P}(y = 1 \mid a, b) = \Lambda(0.1 - 0.1a - 0.5b).$$

Find the marginal effect of increase of b on probability for $a = 0.1$ and $b = 0.4$.

6. Problem

Correlation matrix of three variables is given by

$$C = \begin{pmatrix} 1 & 0.1 & 0 \\ 0.1 & 1 & 0.1 \\ 0 & 0.1 & 1 \end{pmatrix}$$

Eigenvalues with corresponding eigenvectors are

$$\lambda_1 = 1.141, v_1 = \begin{pmatrix} 0.5 \\ 0.707 \\ 0.5 \end{pmatrix}, \lambda_2 = 1, v_2 = \begin{pmatrix} 0.707 \\ 0 \\ -0.707 \end{pmatrix}, \lambda_3 = 0.859, v_3 = \begin{pmatrix} -0.5 \\ 0.707 \\ -0.5 \end{pmatrix}$$

Calculate the change of the first principal component if we increase each original variable by one standard deviation.

7. Problem

Consider 4 points: (6, 6), (10, 6), (6, 5) and (10, 5). Elon Musk applies k-means to find two clusters in these points.

The distance is Euclidean, initial cluster centers are $c_1 = (6, 6)$ and $c_2 = (10, 6)$.

Calculate the center of the second cluster after infinitely many iterations.

8. Problem

Suppose we have the following data:

$$X = \begin{pmatrix} 2 & 2 \\ 2 & 3 \\ 7 & 7 \\ 8 & 7 \\ 10 & 8 \end{pmatrix}$$

We initialize k-means algorithm with initial cluster centers: $c_1 = (4.01, 8.01)$ and $c_2 = (1, 6)$.

Calculate how many observations will be in each class after the first step of the algorithm.

9. Problem

Suppose we have the following distance matrix:

$$\begin{bmatrix} 0 & & & & \\ 2 & 0 & & & \\ 2.8 & 4.5 & 0 & & \\ 2.8 & 2 & 5.7 & 0 & \\ 19.2 & 20.8 & 19.1 & 19.7 & 0 \end{bmatrix}$$

We apply single linkage hierarchical agglomerative clustering to find 2 clusters.

How many observations will be in the biggest cluster?

Single linkage means that the distance between clusters is measured as minimal distance between the members.

10. Problem

On the first step of classic (Torgerson) MDS procedure we have the following distance matrix:

$$\begin{bmatrix} 0 & 6 & 10 \\ 6 & 0 & 15 \\ 10 & 15 & 0 \end{bmatrix}$$

Apply double centering to this matrix.

What will be the value of the element from the second row of first column?

Deadline: 2023-03-26, 21:00.