Chennai Mathematical Institute

Regression and Classification

Final Exam: Aug-Nov, 2023

Total Marks: 80 points
Total Time : 3 hours

Attempt all problems.

Problem 1 (10 points)

a) (1 point) In logistic regression, what distinguishes the probit link function from the logit link function?

b) (2 points) Briefly explain the relationship between cross-entropy and the negative log-likelihood function.

c) (2 points) Why will the performance of linear discriminant analysis and logistic regression be equivalent for the same feature space?

d) (2 points) Define distributed multinomial logistic regression for a k-class classification problem.

e) (3 points) For categorical time series data $\mathcal{D} = y_t : t = 1, 2, \ldots, T$, where $y_t = 0$ with probability p_t and $y_t = 1$ with probability $(1 - p_t)$, develop an auto-regressive time series model of order p within a logistic regression framework.

Problem 2 (10 points)

The linear discriminant rule for class k is presented as:

$$\delta_k(x) = x^T \Sigma^{-1} \mu_k - \frac{1}{2} \mu_k^T \Sigma^{-1} \mu_k + \log \pi_k,$$

the best decision rule is

$$G(x) = \underset{k}{argmax} \ \ \delta_k(x).$$

Consider the following summary statistics from two (independent) data sets:

$$\bar{X}_1 = \begin{pmatrix} 2 \\ 3 \end{pmatrix}, \quad \bar{X}_2 = \begin{pmatrix} 5 \\ 7 \end{pmatrix}, \quad S = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \quad n_1 = 20, \quad n_2 = 40$$

(i) (5 points) Compute the linear discriminant function. Use that linear discriminant function to classify the point $x = [1 \ \ 4]$ as either π_1 or π_2

(ii) (5 points) Under which theoretical assumptions do you expect your method to be reliable?

$\log(2) = 0.693$ and $\log(3) = 1.099$

$(1 \ 4)\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 3 \end{pmatrix} - \frac{1}{2}(2 \ 3)\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 2 \\ 3 \end{pmatrix} - \log 3$

$-\frac{1}{2}(1 \ 1)\begin{pmatrix} 2 \\ 3 \end{pmatrix} - \log 3$

$(-2 \ 3)\begin{pmatrix} 2 \\ 3 \end{pmatrix} - \frac{1}{2}$

$5 - \frac{1}{2} 5 - \log 3 = \frac{5}{2} - \log 3$

$(1 \ 4)\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 5 \\ 7 \end{pmatrix} - \frac{1}{2}(5 \ 7)\begin{pmatrix} 2 & -1 \\ -1 & 1 \end{pmatrix}\begin{pmatrix} 5 \\ 7 \end{pmatrix} + \log 2 - \log 3$

$-\frac{1}{2}(3 \ 2)\begin{pmatrix} 5 \\ 7 \end{pmatrix} + \log 2 - \log 3$

$= (-2 \ 3)\begin{pmatrix} 5 \\ 7 \end{pmatrix} - \frac{1}{2}$

$= 11 - \frac{1}{2} x$

$p = \frac{x}{1 - \ldots}$

$\frac{1-1}{p} = \frac{1+x}{x} - 1$

Problem 3 (10 points)

a) (3 points) Briefly explain when the Gaussian process prior regression model is useful

b) (4 points) Modify the Gaussian process prior regression for a binary classification problem.

c) (3 points) Why does the Gaussian process prior model fail for big data?

Problem 4 (10 points)

a) (3 points) Briefly explain the relationship between basis expansion technique in functional regression and feature engineering in machine learning.

b) (2 points) Briefly explain Kosambi–Karhunen–Loève theorem in the context of Gaussian process prior.

c) (2 points) Explain the relationship between the distance between points and squared-exponential covariance function of Gaussian process model.

d) (3 points) Consider the following statement.

Gauss-Markov Linear Models are special cases of Gaussian Process Regression Models.

Prove or Disprove the above statement.

Problem 5 (10 points)

a) (3 points) What is tree structured regression?

b) (3 points) What is appropraite objective function for tree structured regression? How to fit a tree-structured regression?

c) (4 points) What is intrinsic feature selection? How tree structured regression achieves intrinsic feature selection?

Problem 6 (10 points)

a) (2 points) How non-montonic relationship can be captured in logistic regression?

b) (2 points) Briefly explain the relationship between logistic regression and neural network.

c) (2 points) The solution for Ridge regression is presented as

$$\hat{\beta} = (X^T X + \lambda I)^{-1} X^T y$$

where $\lambda > 0$ is unknown. How can you estimate λ?

d) (4 points) Why Fourier Basis expansion does not suffer from multicollinearity problem?

$(X^T X)^{-1} X^T y = \beta.$

$X^T y = (X^T X)$

$X^T X$ $X^T..$

$X^T X$

Problem 7 (10 points)

Equations of the lines of regression of Y on X is

$$Y - \bar{Y} = r\frac{s_y}{s_x}(X - \bar{X}),$$

and equations of the lines of regression for X on Y are:

$$X - \bar{X} = r\frac{s_x}{s_y}(Y - \bar{Y}).$$

Note that

- $\bar{X}, \bar{Y}$ are the sample mean of X and Y,
- s_x is the sample standard deviation of X
- s_y is the sample standard deviation of Y
- r is the sample correlation coefficients between X and Y.

a) (4 points) If θ is the acute angle between the two regression lines then obtain θ in terms of r, s_x, s_y.

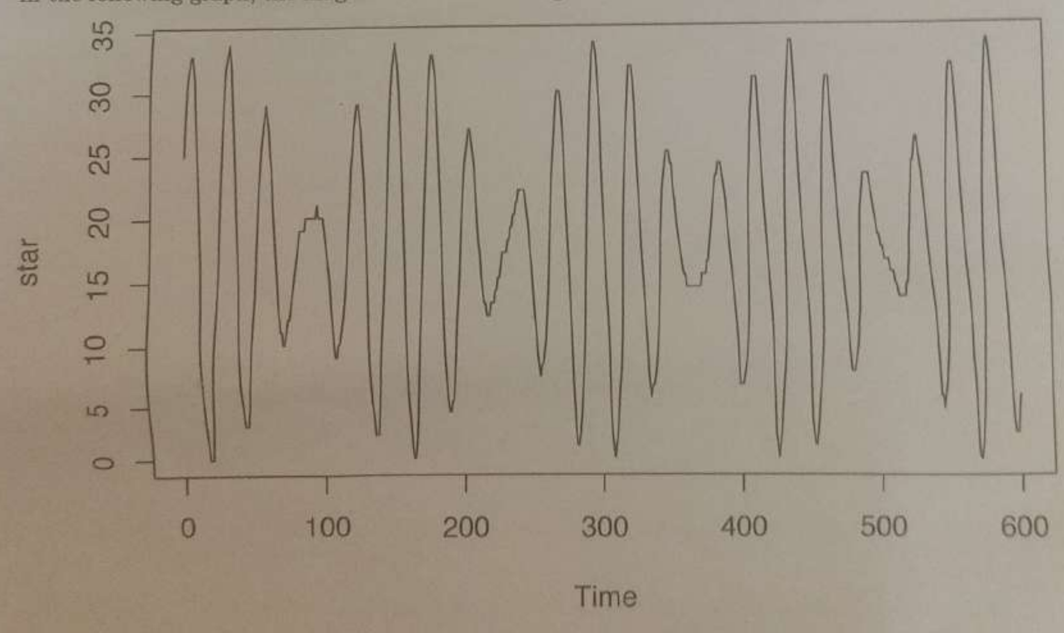
b) (3 points) If $r = 0$ then find θ?

c) (3 points) If $r = 1$ then find θ?

Problem 8 (10 points)

A variable star is a star whose brightness, as observed from Earth, fluctuates over time. These variations can occur for various reasons, leading to the classification of variable stars into different types based on the nature and cause of their brightness changes.

In the following graph, the magnitude of a star's brightness is recorded at midnight for 600 consecutive days.



a) (4 points) Propose an appropriate regression model for predicting the brightness of a star.

b) (4 points) Propose an alternative regression model for consideration.

c) (2 points) What would be your strategy for identifying the best and second-best models?