

DS 501: Statistical and Mathematical Methods for Data Science
Fall 2019
Final Exam Solutions

Problem 1 (Marks: 1+1+4+4)

$z_{0.10} = 1.282$, $z_{0.05} = 1.645$, $z_{0.025} = 1.960$, $z_{0.01} = 2.326$, $z_{0.005} = 2.576$

i. $P(z \geq 1.645) = \underline{0.05}$

ii. $P(-1.96 \leq z \leq 1.645) = \underline{0.925}$

(For part iii and iv) We have to accept or reject the proposal for using a data science software. We take a sample of 200 test points and find 24 misclassifications in it.

iii. Construct a 90% confidence interval for the proportion of misclassifications made by the system. Working required.

Solution

$$0.12 \pm 1.645 \cdot \sqrt{0.12 \cdot 0.88 / 200} = [0.082, 0.158]$$

iv. The developer claims that at the most one in 10 labels are misclassified by the system. Do we have sufficient evidence to disprove this claim at the 5% level of significance?

Working required.

Solution

$$H_0 : p = 1/10$$

$$H_A : \hat{p} > 1/10$$

$$z = (0.12 - 0.10) / \sqrt{0.1 \cdot 0.9 / 200} = 0.94281$$

At 5% significance level the value of z lies in the acceptance region ($z < 1.645$). Hence we have sufficient evidence to support the developer's claim.

Problem 2 (Marks 5+5)

x_1	-1	1	3	-2	-1	1	2	3	1	3
x_2	1	2	3	1	3	-1	1	3	-2	-1
Label	-1	-1	-1	-1	-1	1	1	1	1	1

i. Build a MAP classifier by fitting a 2D-Gaussian distribution to the given set of points. How would you classify the test point (1,2)? Specify all parameters of the classifier. Use unbiased estimates of covariance and variance.

ii. Plot the contours of the distributions for +1 and -1 classes and identify the +1 and -1 regions.

Solution

$$\begin{aligned} \mu_{-1} &= (0, 2) & \Sigma_{-1} &= \begin{bmatrix} 4 & 5/4 \\ 5/4 & 1 \end{bmatrix} \\ \mu_{+1} &= (2, 0) & \Sigma_{+1} &= \begin{bmatrix} 1 & 5/4 \\ 5/4 & 4 \end{bmatrix} \end{aligned}$$

The prior for both classes = 0.5

The classification of (1,2) depends upon the Mahalanobis distance between the test points and the mean of the two classes. Hence it belongs to the negative class according to the given classifier.

Problem 3 (Marks: (4+2) +4)

i. Given the function $f(\mathbf{x}) = x_2 - x_1^2 - 1$ and the constraint $x_2 + 2x_1 + 2 = 0$. Find the stationary points of this function using the method of Lagrange multipliers. Also, draw the contours of this function and the feasible set. Show working.

ii. When training a perceptron, apply one iteration of the stochastic gradient descent algorithm when the momentum is set at 0, learning rate is fixed at $1/4$ and initial weights are $(-1, 1, -2)$ (-1 is the bias). The training point is $(3, 4)$ with the target being $+2$. Show formula and working.

Solution

- i. The stationary point is $(-1, 0)$
 ii. The updated weight is: $(1, 7, 6)$

Problem 4 (Marks: 5)

Find two vectors orthogonal to $[1/\sqrt{6} \quad -1/\sqrt{6} \quad 2/\sqrt{6}]^T$ so that the resulting three vectors form an orthonormal set. Show all working.

Solution

One possible solution:

$$\begin{bmatrix} -2/\sqrt{5} & 0 & 1/\sqrt{5} \end{bmatrix}^T$$

$$\begin{bmatrix} 1/\sqrt{126} & 5/\sqrt{126} & 10/\sqrt{126} \end{bmatrix}^T$$

Problem 5 (Marks: 2 + (1+1+1) +5)

Given the following actual labels and output from a classifier:

Label	+1	-1	-1	-1	+1	+1	-1	-1	+1	+1
Output	0.01	0.2	0.45	0.6	0.65	0.7	0.9	0.95	0.98	0.99

i. Suppose we predict the label as $+1$ if classifier output > 0.5 , then compute:

TP = 4	FP = 3	Precision = $4/7$	Balanced error rate = 0.4
FN = 1	TN = 2	Recall = $4/5$	

ii. Plot the ROC curve for this classifier by computing 5 points on the curve. Write down the computed points and the threshold value for each point.

Solution

5 possible points (FPR, TPR) are:

Threshold = 1	(0,0)
Threshold = .5	(0.6,0.8)
Threshold = 0.25	(0.8,0.8)
Threshold = 0.75	(0.4,0.4)
Threshold = 0	(1,1)

Problem 6 (Marks: 4+3+3)

Consider a hypothetical kitchen with cooks from various cities, wearing three different apron colors, i.e., red, green and blue.

<ul style="list-style-type: none"> • 70% cooks are from Lahore • 20% cooks are from Karachi • 10% cooks are from Multan 	We have the following observations regarding the aprons <ul style="list-style-type: none"> • For Lahori cooks, 30% wear red and 20% wear green aprons • For Karachi cooks, 5% wear red and 45% wear green aprons • For Multani cooks, 10% wear red and 50% wear green aprons
--	---

Use the laws of probability to compute the following. For each part, clearly write down the mathematical formula and show the working clearly. No marks without proper working.

- If we spot a cook wearing a red apron, what is the probability that he is from Lahore?
- If a cook is wearing a green apron, then use maximum likelihood classifier to find the cook's corresponding city.
- If a cook is wearing a blue apron, then use maximum a posteriori classifier to find the cook's city.

Solution

- $P(L|R) = .3 \cdot .7 / (.3 \cdot .7 + .2 \cdot .05 + .1 \cdot .1) = 0.913$
- For Multan has $P(G|M) > P(G|K) > P(G|L)$. Hence Multan is the prediction of the ML classifier
- $P(L|B) = 0.35/P(B)$
 $P(K|B) = 0.10/P(B)$
 $P(M|B) = 0.04/P(B)$

Problem 7 (Marks: 2+4+2+4+2+3+2+2+4)

- Given observations of $x = \{0,0,1,1,1\}$. What is estimate of $P(x=1)$ with Laplacian smoothing?
4/7
- Suppose in a building, there are 30% people who are both short and have brown eyes. There are 60% people who are short. What fraction of short people have brown eyes? Write the rule of probability and show working.
 $P(B|S) = 0.5$

- The L_3 distance between $(1,0,1)$ and $(-1,1,-3) = \underline{\hspace{2cm}} 73^{1/3}$
- Find the singular values and matrix of right singular vectors of: $\begin{bmatrix} 25 & 0 \\ 0 & 16 \end{bmatrix}$ (working not required)

Singular values: 25,16

Matrix of right singular vectors $= \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$

- What is the span of $[1 \ 1 \ 2]^T$ and $[1 \ 1 \ 4]^T$?

All vectors in R^3 of the form (for any value of x and y): $[x \ x \ y]^T$

- Use Chebyshev's inequality to find the bounds on the probability that a point would lie outside the interval $[6,10]$ when the population mean is given by 8 and variance is given by 1. Working required.

$$P(|x-8| \geq 2) \leq 1/4$$

- Do the vectors $[2 \ 2 \ 2]^T$, $[1 \ 1 \ 4]^T$, $[4 \ 4 \ 10]^T$ form an independent set? Give reason for your answer. A simple yes or no is not acceptable.

No

viii. What is the orthogonal projection of $\begin{bmatrix} 1 & -1 & 2 \end{bmatrix}^T$ onto $\begin{bmatrix} 3 & 1 & 4 \end{bmatrix}^T$. Show working.

$$\frac{10}{26} * \begin{bmatrix} 3 & 1 & 4 \end{bmatrix}^T$$

ix. Use least square regression to fit a straight line to the two points (1,3) and (2,0)? Show working and write the final equation of the line. What are the residuals in this case?

$$x^2 = 6 - 3x^1$$

Residuals are both zero