

CS4044D Machine Learning

Monsoon 2021

Assignment 1 (10 marks)

Due Date: 5th, Nov 2021, 10PM

Consider the following table of data.

sample	ω_1			ω_2			ω_3		
	x_1	x_2	x_3	x_1	x_2	x_3	x_1	x_2	x_3
1	-5.01	-8.12	-3.68	-0.91	-0.18	-0.05	5.35	2.26	8.13
2	-5.43	-3.48	-3.54	1.30	-2.06	-3.53	5.12	3.22	-2.66
3	1.08	-5.52	1.66	-7.75	-4.54	-0.95	-1.34	-5.31	-9.87
4	0.86	-3.78	-4.11	-5.47	0.50	3.92	4.48	3.42	5.19
5	-2.67	0.63	7.39	6.14	5.72	-4.85	7.11	2.39	9.21
6	4.94	3.29	2.08	3.60	1.26	4.36	7.17	4.33	-0.98
7	-2.51	2.09	-2.59	5.37	-4.63	-3.65	5.75	3.97	6.65
8	-2.25	-2.13	-6.94	7.18	1.46	-6.66	0.77	0.27	2.41
9	5.56	2.86	-2.26	-7.39	1.17	6.30	0.90	-0.43	-8.71
10	1.03	-3.33	4.33	-7.50	-6.32	-0.31	3.52	-0.36	6.43

1: Write a function (in Python or any language of your choice) to calculate the discriminant function for the given normal density equation (as given below) and prior probabilities. (3 marks)

$$g_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu}_i)^t \boldsymbol{\Sigma}_i^{-1}(\mathbf{x} - \boldsymbol{\mu}_i) - \frac{d}{2} \ln 2\pi - \frac{1}{2} \ln |\boldsymbol{\Sigma}_i| + \ln P(\omega_i).$$

Use your program to complete the following questions.

2: Consider the problem of classifying 10 samples from the above table of data. Assume the that the underlying distributions are normal.

(7 marks)

2. a Assume the prior probabilities of the first two categories are equal and is equal to 1/2 and that of the third category is zero. Design a dichotomizer for those two categories using the feature x_1 alone.

2.b Determine the percentage of points misclassified.

2.c Repeat the above two steps, but now use the two features x_1 and x_2 .

2.d Repeat again, with all the three features taken.

2.e Compare your results and conclude.

2.f Classify the points $(1,2,1)^t$, $(5,3,2)^t$, $(0,0,0)^t$, $(1,0,0)^t$ using each feature vector mentioned above and compare the results.

Submission guidelines

- You can use any programming language.
- You have to submit a single zip file (at the Eduserver) containing two folders: one for each question, with all the related source codes.
- A PDF file containing sample outputs (for different cases of each question) and the design details must also be provided in the zip file.
- Please name your zip file in the format **Firstname_Rollnumber.pdf**