SML Assignment 2, Winter 2022

Deadline: 28th Monday, 11:59 PM

Q1. In this problem, you will explore classification based on discriminant analysis and MLE.[4]

a. Generate 200 multivariate (dimension = 2) Bernoulli distributed samples, 100 of those samples (Class 1) should be generated from (μ1=[0.5 0.8]) and the other 100 (Class2) should be generated from (μ2=[0.9 0.2]). Assume conditional independence. You can use any inbuilt library. .5

b. Select 50 samples out of 100 samples from class 1, call them training set and use those samples to compute the class conditional parameters ⇒ μ1. The remaining 50 samples will be used later for classification, call them test set.

- You need to compute using MLE. You cannot use an inbuilt function that gives MLE.

- Now compute MLE using only n samples at a time where n = 1,2,3,......,50. Analyze MLE vs n plot and report your observations. 1

c. Repeat step (b) for class 2 samples. .5
d. Plot training samples using a scatter plot. .5

e. Derive the discriminant for multivariate Bernoulli and use it for classification for remaining 50 samples of each class. Compute the number of samples correctly classified in each class. You need to show the calculations of discriminant derivation as part of theory and use the same in coding. 1.5

Q2. [Theory] a. Assume a hypothetical prior $\theta^T e^{-\theta}$for d-dimensional multivariate Bernoulli. Derive an expression for θ_{MAP}. This will be in terms of observations. Note, your expression should be a general one and applicable to any d-dimensions and any number of samples. [1.5]

b. Once you derive expression, use following data to find the value. [.5]

$X = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix}$, X is in dxN notation. Note, here θ_{MAP} may come out to be arbitrary real/complex number. Technically it should be a probability. But due to assumption of hypothetical prior, it may not turn out to be probability.

Q1-e)
Multivariate Bernoulli discriminant function
Priors are equal so log on likelihood can be take as disc.
$\ln \prod_{i=1}^{d} \theta_i{}^{x_i}(1-\theta_i)^{1-x_i}$= $\sum_{i=1}^{d} x_i ln\theta_i + (1-x_i)\ln{(1-\theta_i)}$
Theta would come from part b.

$$P(\theta) = \theta^T e^{-\theta}$$

$$P(X|\theta) = \theta^{X}(1-\theta)^{(1-X)} \qquad \theta \in \mathbb{R}^{d\times 1}$$

$$X \in \mathbb{R}^{d\times n}.$$

$$\theta_{MAP} = \underset{\theta}{\text{argmax}}\, P(\theta|X) = \frac{P(X|\theta)\, P(\theta)}{P(X)} \qquad P(x) \text{ is constant}$$

$$P(\theta|X) = P(X|\theta)\, P(\theta)$$

$$\prod_{i=1}^{n}\prod_{j=1}^{d} P(\theta_j|x_{ij}) = \prod_{i=1}^{n}\prod_{j=1}^{d} P(x_{ij}|\theta_j)\, P(\theta_j)$$

$$= \prod_{i=1}^{n}\prod_{j=1}^{d} P(x_{ij}|\theta_j) \cdot \prod_{j=1}^{d} P(\theta_j)$$

$$\prod_{i=1}^{n}\prod_{j=1}^{d} P(\theta_j|x_{ij}) = \prod_{i=1}^{n}\prod_{j=1}^{d} \theta_j^{x_{ij}}(1-\theta_j)^{(1-x_{ij})} \cdot \prod_{j=1}^{d} \theta_j e^{-\theta_j}$$

$$\sum_{i=1}^{n}\sum_{j=1}^{d} \ln P(\theta_j|x_{ij}) = \sum_{i=1}^{n}\sum_{j=1}^{d} x_{ij}\ln\theta_j + (1-x_{ij})\ln(1-\theta_j) + \sum_{j=1}^{d}\ln\theta_j - \sum_{j=1}^{d}\theta_j$$

Now $\dfrac{d\left[\sum_{i=1}^{n}\ln(\theta_j|x_{ij})\right]}{d\theta_j} = 0$

$$\frac{\sum_{i=1}^{n} x_{ij}}{\theta_j} - \frac{n - \sum_{i=1}^{n} x_{ij}}{1-\theta_j} + \frac{1}{\theta_j} - 1 = 0$$

$$\frac{\sum x_{ij}}{\theta_j} - \frac{n - \sum x_{ij}}{1-\theta_j} + \frac{1}{\theta_j} - 1 = 0$$

$$\sum x_{ij} - \cancel{\theta_j \sum x_{ij}} - n\theta_j + \cancel{\theta_j \sum x_{ij}} + 1 - \theta_j - \theta_j + \theta_j^2 = 0$$

$$\sum x_{ij} - n\theta_j - 2\theta_j + \theta_j^2 + 1 = 0$$

$$\theta_j^2 - \theta_j(n+2) + \left(\sum x_{ij} + 1\right) = 0$$

$$\frac{-b \pm \sqrt{b^2 - 4ac}}{2a}.$$

$$\theta = \frac{(n+2) \pm \sqrt{(n+2)^2 - 4\cdot\left(\sum x_{ij}+1\right)}}{2}$$

$$\theta_1 = 3 \pm \sqrt{5}$$

$$\theta_2 = 3 \pm \sqrt{7}.$$

$$x = \begin{bmatrix} 1 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 \end{bmatrix} \quad n=4,\ d=2$$

$$\theta_1 = \frac{(4+2) \pm \sqrt{(4+2)^2 - 4(1+0+1+1)+1}}{2}$$

$$= \frac{6 \pm \sqrt{36-16}}{2}$$

$$= 3 \pm \sqrt{5}$$

$$\theta_2 = \frac{(4+2) \pm \sqrt{(4+2)^2 - 4\cdot(0+1+0+0)+1}}{2}$$

$$= \frac{6 \pm \sqrt{36-8}}{2}$$

$$= \frac{6 \pm \sqrt{28}}{2} = 3 \pm \sqrt{7}..$$

Q3. [Will be continued in Monday' lecture] Take an arbitrary 2x2 matrix. Cannot be the same as in the lecture. Note that the first step should be centralization, that is removal of mean.[4]

a. [Theory] Compute PCA and determine its eigenvectors U and encoding Y = U'centralized(X).
b. [Theory] Find the MSE between UY+mean(X) and X.

parts a,b and c must match.
If not, deduct 0.5 marks

c. For the same matrix, compute (a) and (b) using code. Does your calculation match with the code? You are free to use libraries for eigen decomposition. 1
d. Now take a d-dimensional multivariate Gaussian and sample N realizations. This gives you data matrix X. .5
e. Write a code to compute U for this X. Your code should work for arbitrary d and N. 2
f. Compute MSE between UY + mean(X) and X. Here, Y = U'centralized(X).
g. Plot MSE vs. number of principal components. You need to take first column, first two columns, first three columns and so on till U. Compute MSE for each case between X and U_pY + mean(X), where Y = Up'centralized(X)). 0.5

For part e, you should check for centralization, computation of covariance matrix, eigenvectors and their arrangement in decreasing order of eigenvalues. also see if changing d and N is working.
If changing d or N isnt working, deduct 0.5 marks.

For part g, MSE should decrease with more number of columns of Up and will be least for U.