IE30301-Datamining Assignment 2 (60 Points)

Prof. Junghye Lee

March 30, 2021

Exercise 1

Write the definition and the difference between the below terminologies.[10pts, 2.5pts for each.]

1. **Overfitting** & **Underfitting**
2. **Continuous Value** & **Discrete Value**
3. **Nominal** & **Ordinal**
4. **Univariate** & **Multivariate**

Exercise 2

Consider a multinomial logistic regression model with the dependent variable that has three or more nominal type categories. If we define v_{ij} as the value of category j from the r_i independent trial $\left(\text{instead of the usual binary logistic regression formula } v_{ij} = \begin{cases} 1, & \text{for } y_i = j \\ 0, & \text{for } y_i \neq j \end{cases}\right)$ then the v_{ij} follows a multinomial distribution with probabilities $(P_1, \ldots, P_j)$.
Construct the likelihood function for this case. [10 pts]

Exercise 3

For matrix $\mathbf{A}$, solve the following problems

$$\mathbf{A} = \begin{pmatrix} 8 & -4 \\ 5 & -1 \end{pmatrix}$$

3.1

Compute the eigenvalues λ_1, λ_2 $(\lambda_1 < \lambda_2)$ and its corresponding eigenvectors v_1, v_2 of matrix $\mathbf{A}$ [3 pts]

3.2

Find matrix $\mathbf{P}$ to diagonalize $\mathbf{A}$. Here $\mathbf{D}$ is a diagonal matrix of size 2×2 [3 pts]

$$\mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{D}$$

3.3

Compute the determinant of $\mathbf{A}^4$. The calculation shoud be trivial if you use the properties of determinant. [2 pts]

Exercise 4

Given data $\mathbf{X}$, solving following problems

$$\mathbf{X} = \begin{pmatrix} 1 & 3 & 9 \\ 2 & 5 & 7 \\ 4 & 4 & 6 \\ 9 & 8 & 2 \end{pmatrix}$$

i.e., data with four samples and three features (predictors).

4.1

Find the mean value of each column. [1 pts]

4.2

Subtract each mean from each element of the corresponding column.
Now, let us set the derived matrix as $\mathbf{X}'$. [1 pts]

4.3

Calculate $\frac{1}{4-1}\mathbf{X}'^T\mathbf{X}'$. [1 pts]

4.4

For the calculated matrix in **C**, find eigenvalues $\lambda_1, \lambda_2, \lambda_3$ in descending order up to four decimal places. [2 pts]
(https://www.symbolab.com/solver/matrix-eigenvalues-calculator/eigenvalues)

4.5

Calculate $\frac{\lambda_1}{(\lambda_1+\lambda_2+\lambda_3)}$. [2 pts]

4.6

Find eigenvectors $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ corresponding to $\lambda_1, \lambda_2, \lambda_3$. [2 pts]
(https://www.symbolab.com/solver/matrix-eigenvalues-calculator/eigenvectors)

Exercise 5

Consider three random variables X, Y, Z. The three variables have the covariance matrix in the form of;

$$\Sigma = \begin{pmatrix} \Lambda & \rho\Lambda & 0 \\ \rho\Lambda & \Lambda & \rho\Lambda \\ 0 & \rho\Lambda & \Lambda \end{pmatrix}$$

, where $\frac{1}{\sqrt{2}} < \rho < 1$.
[You may use handwriting only for this question. Please write it down neatly, take a photo and paste it. If the handwriting is illegible, we may not be able to mark this question]

5.1

Calculate the eigenvalues $\lambda_1, \lambda_2, \lambda_3$. (Show process of calculation neatly) [9 pts]

5.2

Find PC(Principal Component) $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ of each random variable X, Y, Z. [9 pts]

5.3

Calculate how much total variance is explained by each principal component. [5 pts]