IE30301-Datamining Assignment 3 (70 Points)

Prof. Junghye Lee

April 11, 2021

Exercise 1

Write a detailed description of the following concepts and their differences. (Explain at least 2 lines about each concepts and write differences in 1 sentence. If not, there are 3 points deduction per problem. [20 pts, 5 pts for each.]

1. **Likelihood** & **Probability**
2. **Feature Selection** & **Feature Extraction**
3. **PC Score** & **PC loading**
4. **Newton-Raphson Method** & **Gradient Descent**

Exercise 2

In multiple linear regression, we can estimate $\hat{\boldsymbol{\beta}}$ as follows by least square method.

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} \in \mathbb{R}^{(p+1)\times 1}, \quad \mathbf{X} \in \mathbb{R}^{N\times(p+1)}, \quad \mathbf{y} \in \mathbb{R}^{N\times 1} \tag{2.1}$$

The regression model can be derived as follows:

$$\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y} = \mathbf{H}\mathbf{y} \tag{2.2}$$

And in the above equation, $\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T$ is specifically referred to as $\mathbf{H}$, hat matrix.

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \tag{2.3}$$

2.1

Show that $\mathbf{H}$ is symmetric ($\mathbf{H}^T = \mathbf{H}$) and idempotent ($\mathbf{H}^2 = \mathbf{H}$). [3 pts]

2.2

An estimator of a given parameter is said to be unbiased if its expected value is equal to the true value of the parameter. ($\mathrm{E}[\hat{\theta}] = \theta$)

Show that $MSE = \hat{\sigma}^2 = \dfrac{SSE}{N-p-1} = \dfrac{(\mathrm{y}-\hat{\mathrm{y}})^T(\mathrm{y}-\hat{\mathrm{y}})}{N-p-1}$ **is unbiased estimator** through using following properties (2.4) - (2.8) & results of above problem **2.1** [10 pts]

If c is a scalar, and **A** is a n × n square matrix, ($c \in \mathbb{R}, \mathbf{A} \in \mathbb{R}^{n\times n}$)

$$c = \mathrm{Tr}(c), \quad \mathrm{Tr}(c\mathbf{A}) = c\mathrm{Tr}(\mathbf{A}) \tag{2.4}$$

If **A, B** are n × n square matrix that have same dimension, ($\mathbf{A}, \mathbf{B} \in \mathbb{R}^{n\times n}$)

$$\mathrm{Tr}(\mathbf{A}+\mathbf{B}) = \mathrm{Tr}(\mathbf{A}) + \mathrm{Tr}(\mathbf{B}), \quad \mathrm{Tr}(\mathbf{A}-\mathbf{B}) = \mathrm{Tr}(\mathbf{A}) - \mathrm{Tr}(\mathbf{B}) \tag{2.5}$$

If **A** is a n × n square matrix, $\mathbf{A} \in \mathbb{R}^{n\times n}$

$$\mathrm{E}[\mathrm{Tr}(\mathbf{A})] = \mathrm{Tr}(\mathrm{E}[\mathbf{A}]) \tag{2.6}$$

If **A** is a n × m matrix and **B** is a m × n matrix ($\mathbf{A} \in \mathbb{R}^{n\times m}, \mathbf{B} \in \mathbb{R}^{m\times n}$)

$$\mathrm{Tr}(\mathbf{AB}) = \mathrm{Tr}(\mathbf{BA}) \tag{2.7}$$

If **x** is random vector,

$$\mathrm{Var}[\mathbf{x}] = \mathrm{E}[\mathbf{x}\mathbf{x}^\mathrm{T}] - \mathrm{E}[\mathbf{x}]\mathrm{E}[\mathbf{x}]^T, \quad \mathrm{E}[\mathbf{x}\mathbf{x}^\mathrm{T}] = \mathrm{Var}[\mathbf{x}] + \mathrm{E}[\mathbf{x}]\mathrm{E}[\mathbf{x}]^T \tag{2.8}$$

Exercise 3

For matrix **A**, solve the following problems

$$\mathbf{A} = \begin{pmatrix} 6 & -3 \\ 5 & -2 \end{pmatrix}$$

3.1

Compute the eigenvalues λ_1, λ_2 ($\lambda_1 < \lambda_2$) and its corresponding eigenvectors v_1, v_2 of matrix **A** [2 pts]

3.2

Find matrix **P** to diagonalize **A**. Here **D** is a diagonal matrix of size 2 × 2 [3 pts]

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

3.3

Compute the determinant of $\mathbf{A}^{2521}$. The calculation shoud be trivial if you use the properties of determinant. [3 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 **4.3**, find eigenvalues $\lambda_1, \lambda_2, \lambda_3$ in descending order ($\lambda_1 > \lambda_2 > \lambda_3$) up to four decimal places. [2 pts]
(https://www.symbolab.com/solver/matrix-eigenvalues-calculator/eigenvalues)

4.5

Calculate $\dfrac{\lambda_1}{(\lambda_1 + \lambda_2 + \lambda_3)}$. (up to four decimal places) [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;

$$\Sigma = \begin{pmatrix} a & ka & 0 \\ ka & a & ka \\ 0 & ka & a \end{pmatrix}$$

, where $0 < k < \frac{1}{\sqrt{2}}$

5.1

Calculate the eigenvalues $\lambda_1, \lambda_2, \lambda_3$. (Show process of calculation neatly) [4 pts]

5.2

Find PC(Principal Component) $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ of each random variable X, Y, Z. [3 pts]

5.3

Calculate how much total variance is explained by each principal component. [3 pts]

Exercise 6

The following table is the outcome of the logistic regression model for an iris flower being species Versicolor versus species Virginica. ($Y = 1$ for Versicolor, $Y = 0$ for Virginica) [9 pts]

Variables	Intercept	Length of Sepals	Width of Sepals	Length of petals	Width of Petals
Estimated Coefficient	25.21	-3	-0.7	2.4	-10.3

6.1

Interpret the effect of length of sepals on the relative risks of an iris flower being species Versicolor versus species Virginica. Fill up the **A,B,C** in the below interpretation of the outcome and select the appropriate word for **D**. (Tips from TA : Think about the meaning of odds(X)!) [4 pts]

Interpretation: Species A is B times more probable than Species C when the length of sepals D(Increase or Decrease) by 1 unit.

x_1	6.4	3.1	4.3	1.3
x_2	6.9	3.0	3.9	1.4

6.2

What is the predicted class for the following new data $\mathbf{x}_1$ and $\mathbf{x}_2$? You should provide the probability of $P(Y_1 = 1|\mathbf{x}_1), P(Y_1 = 0|\mathbf{x}_1), P(Y_2 = 1|\mathbf{x}_2), P(Y_2 = 0|\mathbf{x}_2)$. [6 pts]