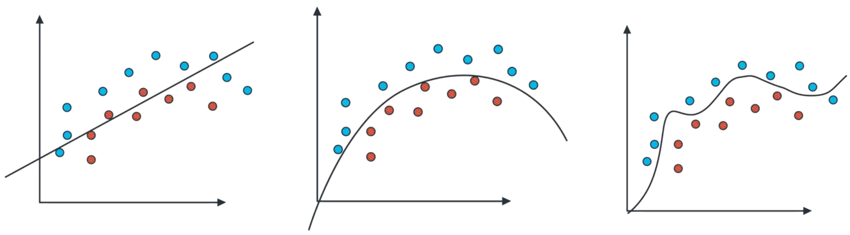
Datamining Assignment 2

Elkhan Ismayilzada

April 17, 2021

Exercise 1

Overfitting & **Underfitting**
Underfitting happens when the machine learning algorithm does not do well in training set and the algorithm overfits when it does great in training set but it tends to memorize it instead of learning characteristics of it. As an example consider following graphs



(a) (b) (c)

Figure 1: (a) Underfitting (b) Good model (c) Overfitting

Given blue and red dots, our job is to separate them using a classification model such as logistic regression. What we see is that, they can easily be separated by using a quadratic curve. If the model underfits then it can draw a line which does not separate dots well. And if the model overfits then it can draw high order polynomial to fit perfectly to the training dataset so that there is no errors.
Continuous Value & **Discrete Value**
Continuous values are the values that can take any value within a range. There exists infinite number of values between two continuous value. Discrete value, on the other hand, can take only certain values that are countable. As an example for discrete values, we can say integers such as 1, 2, 3 etc., and for continuous values, we can say decimals such as 0.68, 0.7 etc.
Nominal & **Ordinal**
Nominal values are the values that are simply "named" or labeled and they have no specified order. Ordinal values, on the other hand, are the values that have some kind of specified order. For ordinal values, days of the week (Monday, Tuesday, Wednesday ...) can be the example. For the nominal values, genders (male / female) , ethnicity (hispanic / asian ...) can be the examples.

Univariate & Multivariate
In Univariate statistics, the algorithm uses only one variable / feature at a time. On the other hand, in multivariate statistics, the algorithm uses two or more than two variables / features.

Exercise 2

In the case of binary logistic regression formula we have

$$\log\left(\frac{P(Y=1)}{1-P(Y=1)}\right) = \beta_0 + \beta x \tag{2.1}$$

However, now we have more than 3 nominal categories, we need to use (2.1) several times for each class. To explain in more detail, suppose we have C number of classes. This means that for each class we need to use (2.1) $(C-1)$ times for each class. Let's say we want to find $P(Y=C)$. The first thing we should do is to consider class C as pivot,

$$\begin{aligned}
\log\left(\frac{P(Y=1)}{P(Y=C)}\right) &= \boldsymbol{\beta}_1^T x \\
\log\left(\frac{P(Y=2)}{P(Y=C)}\right) &= \boldsymbol{\beta}_2^T x \\
\log\left(\frac{P(Y=3)}{P(Y=C)}\right) &= \boldsymbol{\beta}_3^T x \\
\log\left(\frac{P(Y=4)}{P(Y=C)}\right) &= \boldsymbol{\beta}_4^T x \\
&\dots\dots\dots \\
\log\left(\frac{P(Y=C-1)}{P(Y=C)}\right) &= \boldsymbol{\beta}_{c-1}^T x
\end{aligned} \tag{2.2}$$

Notice that from here we can find probabilities of the classes other than class C by exponentiating both sides, so

$$\begin{aligned}
P(Y=1) &= P(Y=C)e^{\beta_1^T x} \\
P(Y=2) &= P(Y=C)e^{\beta_2^T x} \\
P(Y=3) &= P(Y=C)e^{\beta_3^T x} \\
P(Y=4) &= P(Y=C)e^{\beta_4^T x} \\
&\dots\dots\dots \\
P(Y=C-1) &= P(Y=C)e^{\beta_{c-1}^T x}
\end{aligned} \tag{2.3}$$

And if we want to find the $P(Y=C)$, since all C of the probabilities must be one then

$$P(Y=C) = 1 - \sum_{k=1}^{C-1} P(Y=C)e^{\beta_k^T x} = \frac{1}{1+\sum_{k=1}^{C-1} e^{\beta_k^T x}} \tag{2.4}$$

In the same way, other probabilities can also be found as shown here,

$$
\begin{aligned}
P(Y=1) &= \frac{e^{\beta_1^T x}}{1+\sum_{k=1}^{C-1} e^{\beta_k^T x}} \\
P(Y=2) &= \frac{e^{\beta_2^T x}}{1+\sum_{k=1}^{C-1} e^{\beta_k^T x}} \\
P(Y=3) &= \frac{e^{\beta_3^T x}}{1+\sum_{k=1}^{C-1} e^{\beta_k^T x}} \\
P(Y=4) &= \frac{e^{\beta_4^T x}}{1+\sum_{k=1}^{C-1} e^{\beta_k^T x}} \\
&\dots\dots\dots \\
P(Y=C-1) &= \frac{e^{\beta_{c-1}^T x}}{1+\sum_{k=1}^{C-1} e^{\beta_k^T x}}
\end{aligned}
\tag{2.5}
$$

The original equation of multinomial distribution is given as

$$
Pr(Y_{i1}=v_{i1}, Y_{i2}=v_{i2}\dots Y_{iC}=v_{iC}) = \frac{(\sum_{t=1}^{C} v_{it})!}{v_{i1}!\cdot v_{i2}!\cdot\dots v_{iC}!} P(Y_i=1)^{v_{i1}}\cdot P(Y_i=2)^{v_{i2}}\cdot\dots P(Y_i=j)^{v_{iC}} \tag{2.6}
$$

And the maximum likelihood is

$$
L = \prod_{i=1}^{n} \frac{(\sum_{t=1}^{C} v_{it})!}{v_{i1}!\cdot v_{i2}!\cdot\dots v_{iC}!} P(Y_i=1)^{v_{i1}}\cdot P(Y_i=2)^{v_{i2}}\cdot\dots P(Y_i=C)^{v_{iC}} \tag{2.7}
$$

Notice that if we consider v_{ij} as

$$
v_{ij} = \begin{cases} 1, & for \quad y_i = j \\ 0, & for \quad y_i \neq j \end{cases} \tag{2.8}
$$

Then the first term becomes 1. However, in our case v_{ij} follows multinomial distribution with $(P_1, P_2, \dots, P_j)$ So the formula becomes as follows

$$
L = \prod_{i=1}^{n} \frac{(\sum_{t=1}^{C} P_t)!}{P_1!\cdot P_2!\cdot\dots P_j!} P(Y_i=1)^{P_1}\cdot P(Y_i=2)^{P_2}\cdot\dots P(Y_i=C)^{P_C} \tag{2.9}
$$

Exercise 3

3.1

In order to find eigen values we solve following equation

$$
\det\left(\begin{bmatrix} 8 & -4 \\ 5 & -1 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\right) = \det\left(\begin{bmatrix} 8-\lambda & -4 \\ 5 & -1-\lambda \end{bmatrix}\right) = 0 \tag{3.1}
$$

Which is at the end equal to

$$
\det\left(\begin{bmatrix} 8-\lambda & -4 \\ 5 & -1-\lambda \end{bmatrix}\right) = \lambda^2 - 7\lambda + 12 = 0 \tag{3.2}
$$

After solving the equation we get these values

$$\lambda_1 = 4, \lambda_2 = 3 \tag{3.3}$$

After finding eigen values we can easily calculate eigen vectors by putting eigen values in (3.1) without determinant as follows

$$\begin{gathered}
\begin{bmatrix} 4 & -4 \\ 5 & -5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad for \quad \lambda = 4 \\
\begin{cases} 4x - 4y = 0 \\ 5x - 5y = 0 \end{cases} \\
\begin{cases} 4x = 4y \\ 5x = 5y \end{cases} \\
\begin{cases} x = y \\ x = y \end{cases}
\end{gathered} \tag{3.4}$$

If we choose y as 1 then

$$v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{3.5}$$

In the same way we can find the second eigen vector

$$\begin{gathered}
\begin{bmatrix} 5 & -4 \\ 5 & -4 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \quad for \quad \lambda = 3 \\
\begin{cases} 5x - 4y = 0 \\ 5x - 4y = 0 \end{cases} \\
\begin{cases} 5x = 4y \\ 5x = 4y \end{cases} \\
\begin{cases} x = \frac{4}{5}y \\ x = \frac{4}{5}y \end{cases}
\end{gathered} \tag{3.6}$$

So if we choose y as 5 then

$$v_2 = \begin{bmatrix} 4 \\ 5 \end{bmatrix} \tag{3.7}$$

3.2

Diagonal matrix D is composed of eigen values as below

$$D = \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix} \tag{3.8}$$

Now we can find P since it is the only unknown matrix in the equation

$$P^{-1} \begin{bmatrix} 8 & -4 \\ 5 & -1 \end{bmatrix} P = \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix} \tag{3.9}$$

After tedious calculation we find that

$$P = \begin{bmatrix} 1 & 4 \\ 1 & 5 \end{bmatrix} \tag{3.10}$$

3.3

We can find power of the matrix by simply using (3.9)

$$A^4 = PD^4P^{-1} = \begin{bmatrix} 1 & 4 \\ 1 & 5 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 3 \end{bmatrix}^4 \begin{bmatrix} 5 & -4 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 956 & -700 \\ 875 & -619 \end{bmatrix} \tag{3.11}$$

Now we can simply find the determinant of the matrix

$$\det\left(\begin{bmatrix} 956 & -700 \\ 875 & -619 \end{bmatrix}\right) = 956 \cdot (-619) - 875 \cdot (-700) = 20736 \tag{3.12}$$

Exercise 4

4.1

The mean of each column is following matrix

$$mean(X) = \begin{bmatrix} 4 & 5 & 6 \end{bmatrix} \tag{4.1}$$

4.2

To substract each element from mean we should broadcast the matrix as like this

$$X' = \begin{bmatrix} 1 & 3 & 9 \\ 2 & 5 & 7 \\ 4 & 4 & 6 \\ 9 & 8 & 2 \end{bmatrix} - \begin{bmatrix} 4 & 5 & 6 \\ 4 & 5 & 6 \\ 4 & 5 & 6 \\ 4 & 5 & 6 \end{bmatrix} = \begin{bmatrix} -3 & -2 & 3 \\ -2 & 0 & 1 \\ 0 & -1 & 0 \\ 5 & 3 & -4 \end{bmatrix} \tag{4.2}$$

4.3

After putting values

$$\frac{1}{4-1} \begin{bmatrix} -3 & -2 & 0 & 5 \\ -2 & 0 & -1 & 3 \\ 3 & 1 & 0 & -4 \end{bmatrix} \begin{bmatrix} -3 & -2 & 3 \\ -2 & 0 & 1 \\ 0 & -1 & 0 \\ 5 & 3 & -4 \end{bmatrix} = \begin{bmatrix} 38/3 & 7 & -31/3 \\ 7 & 14/3 & -6 \\ -31/3 & -6 & 26/3 \end{bmatrix} \tag{4.3}$$

4.4

To find the eigen values, we use following equation (det stands for determinant)

$$\det\left(\begin{bmatrix} 38/3 & 7 & -31/3 \\ 7 & 14/3 & -6 \\ -31/3 & -6 & 26/3 \end{bmatrix} - \lambda \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right) = \det\left(\begin{bmatrix} 38/3 - \lambda & 7 & -31/3 \\ 7 & 14/3 - \lambda & -6 \\ -31/3 & -6 & 26/3 - \lambda \end{bmatrix}\right) = 0 \tag{4.4}$$

Which is at the end equal to

$$\det\left(\begin{bmatrix} 38/3 - \lambda & 7 & -31/3 \\ 7 & 14/3 - \lambda & -6 \\ -31/3 & -6 & 26/3 - \lambda \end{bmatrix}\right) = \frac{-9\lambda^3 + 234\lambda^2 - 158\lambda - 3894}{9} + 434 = 0 \tag{4.5}$$

After solving the equation we find the values as follows

$$\lambda_1 = 25.3084, \lambda_2 = 0.6044, \lambda_3 = 0.0872 \tag{4.6}$$

4.5

After putting in the values

$$\frac{25.3084}{25.3084 + 0.6044 + 0.0872} = 0.9734 \tag{4.7}$$

4.6

After finding eigen values we can easily calculate eigen vectors by putting eigen values in (5.3) without determinant as follows

$$\begin{bmatrix} -12.6418 & 7 & -31/3 \\ 7 & -20.6418 & -6 \\ -31/3 & -6 & -16.6418 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad for \quad \lambda = 25.3084 \tag{4.8}$$

After solving the equation we get

$$a_1 = \begin{bmatrix} -1.2045 \\ -0.6992 \\ 1 \end{bmatrix} \tag{4.9}$$

In the same we can find second eigen vector

$$\begin{bmatrix} 12.0622 & 7 & -31/3 \\ 7 & 4.0622 & -6 \\ -31/3 & -6 & 8.0622 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad for \quad \lambda = 0.6044 \tag{4.10}$$

Hence we get

$$a_2 = \begin{bmatrix} 138.5863 \\ -237.3327 \\ 1 \end{bmatrix} \tag{4.11}$$

Finally, we go for the last eigen vector as follows

$$\begin{bmatrix} 12.5795 & 7 & -31/3 \\ 7 & 4.5795 & -6 \\ -31/3 & -6 & 8.5795 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad for \quad \lambda = 0.0872 \tag{4.12}$$

Hence we get

$$a_3 = \begin{bmatrix} 0.6182 \\ 0.3652 \\ 1 \end{bmatrix} \tag{4.13}$$

Exercise 5

5.1

To find eigenvalues, we should solve following equation

$$\begin{aligned}
\det\left(\begin{bmatrix} \Lambda & \rho\Lambda & 0 \\ \rho\Lambda & \Lambda & \rho\Lambda \\ 0 & \rho\Lambda & \Lambda \end{bmatrix} - \lambda\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\right) = \det\left(\begin{bmatrix} \Lambda - \lambda & \rho\Lambda & 0 \\ \rho\Lambda & \Lambda - \lambda & \rho\Lambda \\ 0 & \rho\Lambda & \Lambda - \lambda \end{bmatrix}\right) &= 0 \\
(\Lambda - \lambda)([\Lambda - \lambda]^2 - \rho^2\Lambda^2) - \rho^2\Lambda^2(\Lambda - \lambda) &= 0 \\
(\Lambda - \lambda)^3 - (\Lambda - \lambda)\rho^2\Lambda^2 - \rho^2\Lambda^2(\Lambda - \lambda) &= 0 \\
(\Lambda - \lambda)^3 - 2(\Lambda - \lambda)\rho^2\Lambda^2 &= 0 \\
(\Lambda - \lambda)([\Lambda - \lambda]^2 - 2\rho^2\Lambda^2) &= 0 \\
(\Lambda - \lambda)(\Lambda - \lambda - \sqrt{2}\rho\Lambda)(\Lambda - \lambda + \sqrt{2}\rho\Lambda) &= 0
\end{aligned} \tag{5.1}$$

So in the end we have following eigen values

$$\lambda_1 = \Lambda, \lambda_2 = \Lambda - \sqrt{2}\rho\Lambda, \lambda_3 = \Lambda + \sqrt{2}\rho\Lambda, \tag{5.2}$$

5.2

Principal components mean we should find its eigen vectors by using following equation

$$\begin{gathered}
\begin{bmatrix} \Lambda - \Lambda & \rho\Lambda & 0 \\ \rho\Lambda & \Lambda - \Lambda & \rho\Lambda \\ 0 & \rho\Lambda & \Lambda - \Lambda \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad for \quad \lambda = \Lambda \\
\begin{bmatrix} 0 & \rho\Lambda & 0 \\ \rho\Lambda & 0 & \rho\Lambda \\ 0 & \rho\Lambda & 0 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad for \quad \lambda = \Lambda \\
\begin{cases} \rho\Lambda y = 0 \\ \rho\Lambda x + \rho\Lambda z = 0 \\ \rho\Lambda y = 0 \end{cases} \\
\begin{cases} y = 0 \\ x = -z \end{cases} \\
y = 0, x = 1, z = -1
\end{gathered} \tag{5.3}$$

So in the end we have

$$p_1 = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \tag{5.4}$$

In the same way, we find out another principal component

$$\begin{bmatrix} \Lambda-\Lambda+\sqrt{2}\rho\Lambda & \rho\Lambda & 0 \\ \rho\Lambda & \Lambda-\Lambda+\sqrt{2}\rho\Lambda & \rho\Lambda \\ 0 & \rho\Lambda & \Lambda-\Lambda+\sqrt{2}\rho\Lambda \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad for \quad \lambda = \Lambda - \sqrt{2}\rho\Lambda$$

$$\begin{bmatrix} \sqrt{2}\rho\Lambda & \rho\Lambda & 0 \\ \rho\Lambda & \sqrt{2}\rho\Lambda & \rho\Lambda \\ 0 & \rho\Lambda & \sqrt{2}\rho\Lambda \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad for \quad \lambda = \Lambda - \sqrt{2}\rho\Lambda$$

$$\begin{cases} \sqrt{2}\rho\Lambda x + \rho\Lambda y = 0 \\ \rho\Lambda x + \sqrt{2}\rho\Lambda y + \rho\Lambda z = 0 \\ \rho\Lambda y + \sqrt{2}\rho\Lambda z = 0 \end{cases}$$

$$\begin{cases} \rho\Lambda x = -\frac{\rho\Lambda y}{\sqrt{2}} \\ -\frac{\rho\Lambda y}{\sqrt{2}} + \sqrt{2}\rho\Lambda y - \frac{\rho\Lambda y}{\sqrt{2}} = 0 \\ \rho\Lambda z = -\frac{\rho\Lambda y}{\sqrt{2}} \end{cases} \tag{5.5}$$

$$\begin{cases} x = -\frac{y}{\sqrt{2}} \\ -\frac{y}{\sqrt{2}} + \sqrt{2}y - \frac{y}{\sqrt{2}} = 0 \\ z = -\frac{y}{\sqrt{2}} \end{cases}$$

$$\begin{cases} x = -\frac{y}{\sqrt{2}} \\ -2y + 2y = 0 \\ z = -\frac{y}{\sqrt{2}} \end{cases}$$

$$\begin{cases} x = -\frac{y}{\sqrt{2}} \\ 0 = 0 \\ z = -\frac{y}{\sqrt{2}} \end{cases}$$

Let's choose y as $(-\sqrt{2})$ for easy eigen vector then the second principal component is

$$p_2 = \begin{bmatrix} 1 \\ -\sqrt{2} \\ 1 \end{bmatrix} \tag{5.6}$$

Now we are going to find our last principal component

$$\begin{bmatrix} \Lambda - \Lambda - \sqrt{2}\rho\Lambda & \rho\Lambda & 0 \\ \rho\Lambda & \Lambda - \Lambda - \sqrt{2}\rho\Lambda & \rho\Lambda \\ 0 & \rho\Lambda & \Lambda - \Lambda - \sqrt{2}\rho\Lambda \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad for \quad \lambda = \Lambda + \sqrt{2}\rho\Lambda$$

$$\begin{bmatrix} -\sqrt{2}\rho\Lambda & \rho\Lambda & 0 \\ \rho\Lambda & -\sqrt{2}\rho\Lambda & \rho\Lambda \\ 0 & \rho\Lambda & -\sqrt{2}\rho\Lambda \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \quad for \quad \lambda = \Lambda + \sqrt{2}\rho\Lambda$$

$$\begin{cases} -\sqrt{2}\rho\Lambda x + \rho\Lambda y = 0 \\ \rho\Lambda x - \sqrt{2}\rho\Lambda y + \rho\Lambda z = 0 \\ \rho\Lambda y - \sqrt{2}\rho\Lambda z = 0 \end{cases}$$

$$\begin{cases} \rho\Lambda x = \frac{\rho\Lambda y}{\sqrt{2}} \\ \frac{\rho\Lambda y}{\sqrt{2}} - \sqrt{2}\rho\Lambda y + \frac{\rho\Lambda y}{\sqrt{2}} = 0 \\ \rho\Lambda z = \frac{\rho\Lambda y}{\sqrt{2}} \end{cases} \tag{5.7}$$

$$\begin{cases} x = \frac{y}{\sqrt{2}} \\ \frac{y}{\sqrt{2}} - \sqrt{2}y + \frac{y}{\sqrt{2}} = 0 \\ z = \frac{y}{\sqrt{2}} \end{cases}$$

$$\begin{cases} x = -\frac{y}{\sqrt{2}} \\ 2y - 2y = 0 \\ z = -\frac{y}{\sqrt{2}} \end{cases}$$

$$\begin{cases} x = \frac{y}{\sqrt{2}} \\ 0 = 0 \\ z = \frac{y}{\sqrt{2}} \end{cases}$$

Let's choose y as $\sqrt{2}$ for easy eigen vector then the last principal component is

$$p_3 = \begin{bmatrix} 1 \\ \sqrt{2} \\ 1 \end{bmatrix} \tag{5.8}$$

5.3

Total variance is the sum of the diagonals of the variance-covariance matrix which is in our case 3Λ. Explained variance for each principal component can be found as follows:

$$\begin{aligned} p_1(\lambda = \Lambda) &= \frac{\Lambda}{3\Lambda} = 0.3333\cdots \approx 33\% \\ p_2(\lambda = \Lambda - \sqrt{2}\rho\Lambda) &= \frac{\Lambda - \sqrt{2}\rho\Lambda}{3\Lambda} = \frac{1}{3} - \frac{\sqrt{2}}{3}\rho \quad for \quad 0 \le \rho < \frac{1}{\sqrt{2}} \\ p_3(\lambda = \Lambda + \sqrt{2}\rho\Lambda) &= \frac{\Lambda + \sqrt{2}\rho\Lambda}{3\Lambda} = \frac{1}{3} + \frac{\sqrt{2}}{3}\rho \quad for \quad 0 \le \rho < \frac{1}{\sqrt{2}} \end{aligned} \tag{5.9}$$