2 Mathematical Methods

2.1 Vectors

The first math concept is called vectors, a column of n numbers. If $\mathbf{x}$ is a vector, then $\mathbf{x}$ is in the form of

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

where $x_1, ..., x_n$ belongs to real numbers. n is called the number of components of $\mathbf{x}$, and correspondingly, $\mathbf{x}$ is an n-vector. The length of an n-vector $\mathbf{x}$ is $\|\mathbf{x}\| = \sqrt{x_1^2 + ... + x_n^2}$. The distance between two n-vectors, $\mathbf{x}$ and $\mathbf{y}$ is the norm of the difference $\text{dist}(\mathbf{x}, \mathbf{y}) = \|\mathbf{y} - \mathbf{x}\|$

2.2 Linear Regression

For each point given, it is feasible to write the point in the form of $\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} ... \begin{bmatrix} x_n \\ y_n \end{bmatrix}$. Then we can construct two n-vectors, $\mathbf{x}$ and $\mathbf{y}$ respectively, to represent the collections of $x_1, ..., x_n$ and $y_1, ..., y_n$.

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ ... \\ x_n \end{bmatrix} \text{ and } \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{bmatrix}$$

The best fit line can be expressed in the form of $y = \alpha x + \beta$; accordingly, we can construct an n-vector whose values are the y values of this line corresponding to each x.

$$\begin{bmatrix} \alpha x_1 + \beta \\ \alpha x_2 + \beta \\ ... \\ \alpha x_n + \beta \end{bmatrix}$$

Since the line best fits the data, α and β should be the numbers that make the distance between $\mathbf{y}$ and the vector of the function $y = \alpha x + \beta$ as small as possible.

Assume that the function J is defined as the average of the sum of the square of the distance from y to $\alpha x + \beta$, so it can be expressed in the following equation.

$$J(\alpha, \beta) = \frac{1}{n} \sum_{i=1}^{n} (\alpha x_i + \beta - y_i)^2$$

The reason why the function needs to be squared is that the accumulative difference will not compensated by the negative difference if the total difference is positive. Conversely, the positive difference will compensate the total difference if the sum of the difference is negative. As a result, when the function is squared, the accumulative difference can only increase, which is convenient for finding the best fit line.

From the function J, there are two observations: The first one is that the α and β that best fits the data is the α and β that minimize J. It is because the function is an expression that represents the total distance between the best fit line and the original points; shorter distance means the line fits the data points better. The second one is that the function is in a quadratic form. Every quadratic function can be written in the form of $ax^2 + bx + c$. In the function J, the coefficient before x^2, a, is larger than 0 because α^2 is always positive. Therefore, the function $ax^2 + bx + c$ attains its minimum value when $x = -\frac{b}{2a}$.

If β is regarded as a constant and α is the variable, after expanding the function J, the formula can be written in the following form:

$$J = \frac{1}{n}\Big(\sum_{i=1}^{n} x_i^2\Big)\alpha^2 + \frac{2}{n}\Big(\sum_{i=1}^{n}(\beta - y_i)x_i\Big)\alpha + \frac{1}{n}\sum_{i=1}^{n}(\beta - y_i)^2$$

The second observation indicates that J can attain its minimum value when

$$\alpha = -\frac{\sum_{i=1}^{n}(\beta - y_i)x_i}{\sum_{i=1}^{n} x_i^2}$$

Likewise, If α is regarded as a constant and β is the variable, the function can be expressed in the following form:

$$J = \beta^2 + \frac{2}{n}\Big(\sum_{i=1}^{n}(\alpha x_i - y_i)\Big)\beta + \frac{1}{n}\sum_{i=1}^{n}(\alpha x_i - y_i)^2$$

Now, J can reach its lowest point when

$$\beta = -\frac{1}{n}\sum_{i=1}^{n}(ax_i - y_i)$$

After algebraic manipulations of the equations of the minimum α and β, function J can be minimized with the α and β satisfying the following system of equations:

$$\Big(\sum_{i=1}^{n} x_i^2\Big)\alpha + \Big(\sum_{i=1}^{n} x_i\Big)\beta = \sum_{i=1}^{n} x_i y_i$$

$$\Big(\sum_{i=1}^{n} x_i\Big)\alpha + n\beta = \sum_{i=1}^{n} y_i$$

In order to solve the system of linear equations of two unknowns, α and β in this case, the solution formula for equations with two variables can be recalled. Suppose the system of equations are

$$a_{11}\alpha + a_{12}\beta = b_1$$

$$a_{21}\alpha + a_{22}\beta = b_2$$

The system of equations has exactly one solution if and only if

$$\Delta = a_{11}a_{22} - a_{12}a_{21} \neq 0$$

Then the solution will be

$$\alpha = \frac{a_{22}b_1 - a_{12}b_2}{\Delta}$$

$$\beta = \frac{a_{11}b_2 - a_{21}b_1}{\Delta}$$

Since this solution concludes the mathematical methods part, an example of the English Premier League will be presented in the next section.