Linear Regression

Regresyon katsayılarını bulmak için 2 yöntem kullanılır

1) Ordinary Least Square:
   - Non iterative Method
   - Analytical Solution
   - Not Scalable

2) Gradient Descent:
   - Iterative Method
   - Optimizasyon Algoritması
   - Used for minimising Cost Function
   - Birçok şeyi optimize etmek için kullanılır

→ Daha öncesinde Gradient Descent'i nasıl kullanacağımızı gördük. Bu analiz OLS ile yapılacak olup; Gradient Descent en sonda ayrıca çok kısa bir şekilde anlatılacaktır.

1) Simple Linear Regression

$e_i = y_i - \hat{y}_i$; Sum of Squared residuals $= e_1^2 + e_2^2 + \dots + e_n^2$

$$= (y_1 - \hat{\beta}_0 - \hat{\beta}_1 x_1)^2 + (y_2 - \hat{\beta}_0 - \hat{\beta}_1 x_2)^2 + \dots + (y_n - \hat{\beta}_0 - \hat{\beta}_1 x_n)^2$$

The OLS approach chooses $\hat{\beta}_0$ and $\hat{\beta}_1$ to minimize the our Loss Function (Sum of Squared Residuals), Using some calculus it is shown that the minimizers are:

$$\underbrace{\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \qquad \hat{\beta}_1}_{\text{OLS Solution}} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2} = \frac{\text{Covariance}}{\text{Variance}}$$

We can find how close $\hat{\beta}_0$ and $\hat{\beta}_1$ are to the True Values β_0 and β_1. To compute the Standard Errors associated with $\hat{\beta}_0$ and $\hat{\beta}_1$, we use the following formulas:

↓

$$SE(\hat{\beta}_0)^2 = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right] ; \quad SE(\hat{\beta}_1)^2 = \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

- Where $\sigma^2 = Var(\varepsilon)$.

→ Standard Errors can used to perform Hypothesis Tests on the Coefficients.

$H_0 \Rightarrow \beta_1 = 0$

$H_1 \Rightarrow \beta_1 \neq 0$

↳ Değişkenin tek başına anlamlı olup veya olmadığını ölçeriz.

if $\beta_1 = 0$ then the model becomes $y_i = \beta_0 + \varepsilon$

$$t = \frac{\hat{\beta}_1 - 0}{SE(\hat{\beta}_1)}$$ ⇒ β_1'in 0'dan kaç standard sapma uzak olduğunu ölçer.

t value ile p-value'nun çok yakın bir ilişkisi vardır.

$\beta_1 = 0$ farz edildiğinde bulunan t değerine eşit veya daha büyük herhangi bir değer gözlemleme olasılığına *p-value diyoruz.

↓

* t değeri arttıkça β_1'in 0'dan uzaklaştığını biliyoruz.

* p value ise bize şunu diyor: $\beta_1 = 0$ iken bulunan "t" değerine eşit veya daha büyük bir t gözlemleme olasılığı. Ama biz $\beta_1 = 0$ iken "t" değerinin çok küçük olmasını bekleriz. Yani "p olasılığı"; "t" büyük iken $\beta_1 = 0$ olma olasılığını verir.

Simple Linear Regression Problemini nasıl çözebiliriz:

1) OLS (Analytical Solution)
2) Least Squares Normal Equation (Matris)
3) Optimization Algorithm (ex: Gradient Des)

2) Multiple Linear Regression

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_p x_p$$

Sum of Squared Residuals

The values $\hat{\beta}_0, \hat{\beta}_1, \dots, \hat{\beta}_p$ that minimizes the **Loss Function** are the multiple least squares regression coefficient estimates

Multiple Linear Regression
Problemini nasıl çözebiliriz:

1) OLS (Analytical Solution)

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2$$

$$\hat{\beta}_1 = \frac{(\sum x_2^2)(\sum x_1 y) - (\sum x_1 x_2)(\sum x_2 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

$$\hat{\beta}_2 = \frac{(\sum x_1^2)(\sum x_2 y) - (\sum x_1 x_2)(\sum x_1 y)}{(\sum x_1^2)(\sum x_2^2) - (\sum x_1 x_2)^2}$$

Görüldüğü gibi 2 bağımsız değişkenli modelde bile çok fazla işlem var. Bu complex işlemlerden kurtulmak için problemi Matrix yaklaşımı ile çözeriz.

2) Least Squares Normal Equation (Matrix Form)

The Least Squares Normal Equation can be expressed as:

$$\boxed{X^T X b = X^T Y}$$

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \quad b = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_n \end{bmatrix}$$

$$\hat{y} = b_0 + \hat{b}_1 x_1 + \dots + \hat{b}_k x_k$$

$$X = \begin{bmatrix} 1 & X_{1,1} & X_{1,2} & \cdots & X_{1,k} \\ 1 & X_{2,1} & X_{2,2} & \cdots & X_{2,k} \\ \vdots & & & & \\ 1 & X_{n,1} & X_{n,2} & \cdots & X_{n,k} \end{bmatrix}$$

3) Optimizasyon Algoritması (ex: Gradient Descent)

Daha önce anlatılmıştı.

Devam edelim;

Model overall anlamlılığı:

$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$

H_1: at least one β_j is non-zero.

(F istatistiğinin de büyük olmasını isteriz)

$$F\text{-statistic} = \frac{\text{Açıklanan Varyans}}{\text{Açıklanamayan Varyans}} = \frac{SSR/k}{SSE/n-k-1}$$

→ P-value which is associated with F statistic is calculated. Based on this p-value, we can determine whether accept or reject H_0.

↓

(Nasıl ki "t istatistiğini" p-value ile yorumladık, aynı yorumu tamamen "F istatistiği" için de yapabiliriz.)

Gradient Descent üzerine;

Coefficient'ları iterative optimize eden ve amacı Cost Function'ı minimize etmek olan algoritmadır.

(Nasıl çalıştığı ile ilgili detayı Gradient Descent bölümünde vermiştik)

↳ Each iteration updates $\beta_0, \beta_1, \dots, \beta_n$ to a line that yields slightly lower error than the previous iteration.