

Udemy - Vahit Keskin - (10 Saat) Python A-Z: Veri Bilimi ve M.L

260. Basit Doğrusal Regresyon - Teori

Anahtar teorik gösterim: $y = \beta_0 + \beta_1 x + \epsilon$

Örnekleme gerçek değerler: $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$

Tahmin modeli: $\hat{y}_i = \beta_0 + \beta_1 x_i$

β_0 = Doğrunun y eksenini kestiği nokta

β_1 = Doğrunun eğimi

ϵ = Hata termi

Basit doğrusal regresyon olduğu için sadece 1 tane bağımlı (x) değişkenimiz var.

Amacımız β_0 ve β_1 'i (β_0 ve β_1 'in tahminleri) bulmak.

Donanımızda bulmuş olduğumuz β_0 ve β_1 'yi $\hat{y}_i = \beta_0 + \beta_1 x_i$ şeklinde denkleme oturtup bir tahmin modeli haline getiriyoruz (bunlar öğrenilmiş kat sayılar oluyor artık)

Örnekleme teorik gösterim:

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

$$\epsilon_i = y_i - \beta_0 + \beta_1 x_i$$

(- \hat{y}_i)

Tahmin modeli:

$$SSE = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

$$\hat{y}_i = \beta_0 + \beta_1 x_i$$

Hatalar /artıklar:

$$SSE = \sum_{i=1}^n (y_i - \underbrace{(\beta_0 + \beta_1 x_i)}_{\hat{y}_i})^2$$

$$\epsilon_i = y_i - \hat{y}_i$$

SSE: Sum of Squares Error
Hata kareler toplamı

SSE denklemini çözerek β_0 ve β_1 kat sayılarını bulmak olacak amacımız. Bu denklemi çözmek için β_0 ve β_1 'e göre türev almalıyız gerekiyor. Türev aldığımızda β_0 ve β_1 'i açık dökümler buluyoruz.

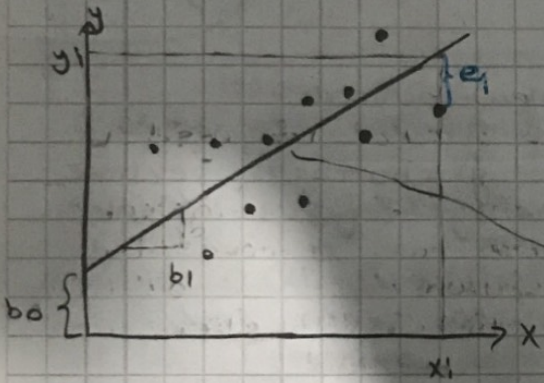
$$SSE = \sum_{i=1}^n (y_i - (b_0 + b_1 x_i))^2 \quad \hat{y}_i$$

$$b_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad b_0 = \bar{y} - b_1 \bar{x}$$

\bar{x} : Bağımsız değişkenin ortalaması

\bar{y} : Bağımlı değişkenin ortalaması

Basit Doğrusal Regresyon Geometrik Gösterim



Nokta \rightarrow gerçek y değerleri

$$y_i = b_0 + b_1 x_i + e_i$$

Gerçek y değerlerine bu formülle erişiyoruz

Tahmin fonksiyonu ve tahmin edilen değerler

$$\hat{y}_i = b_0 + b_1 x_i$$

Yukarıdaki b_0 ve b_1 formüllerini uyguladığımızda bulduğumuz b_0 ve b_1 kat sayılarıyla $\hat{y}_i = b_0 + b_1 x_i$ tahmin fonksiyonumuzu oluşturuyoruz.