

337. Lojistik Regresyon - Teori

Amaç sınıflandırma problemi için bağımlı ve bağımsız değişkenler arasındaki ilişkiyi tanımlayan doğrusal bir model kurmaktır. Burada y bağımlı değişkenimiz kategoriktir. x 'in her bir kategoriye ait olma olasılıklarıyla ilgilenir. Sınıf tahminlerinin yanlış yönlü sınıflandırma oranı ya da doğru sınıflandırma oranıyla ölçülür. Lineer regresyon yerine lojistik regresyon kullanıyoruz (Nedensizlik kitabında daha güzel açıklanmıştır) çünkü lojistik regresyonda kullanılan fonksiyondan dolayı tahmin edilen değerlerin 0 ile 1 arasında olması durumu garanti altına alınmış olur ve ilgilendirmiş sınıfta 1 olması ile ilgili bir olasılık değeri hesaplanmış olur. Belirtilecek olan bir eşik değere göre olasılık değerlerini değerlendiriyoruz.

Lojistik Dağılım Fonksiyonu

$$\pi(x) = P(Y=1|X=x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p}}$$

Burada x bağımsız değişkenleri değerlerini aldığında y bağımlı değişkeninin 1. sınıfa ait olması olasılığı bu şekilde hesaplanıyor.

Lojistik Dağılım Fonksiyonunun Odds'u

$$\frac{\pi(x)}{1 - \pi(x)}$$

Lojistik Regresyon Modeli
Oddsun logaritmasını alıyoruz.

$$g(x) = \ln \frac{\pi(x)}{1 - \pi(x)} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p$$

Yine doğrusal bir yapı elde ettik. Neden doğruluk oluyor? Yorumlanabilirlikten dolayı.

- Bağımlı değişken kategoriktir.
- Adını bağımlı değişkene uygulanan logit dönüşümünden alır.
- Doğrusal Regresyonda aranan varsayımlar burada aranmadığı için daha esnek kullanılabilirliği vardır.
- Bağımlı değişkenin 1 olarak tanımlanan değerinin gerçekleşme olasılığı hesaplanır. Böylelikle bağımlı değişkenin alacağı değer ile karşılaştırılır.
- Logistik fonksiyonu sayesinde üretilen değerler 0-1 arasında olur.