

Doğrusal Regresyon

Şevval Murat

Doğrusal regresyon, **bağımlı değişken** (y) ile **bağımsız değişkenler** (x) arasındaki doğrusal ilişkiyi modellemek için kullanılan temel bir istatistiksel yöntemdir. Örneğin, **ev fiyatlarını tahmin etmek** için oda sayısı ve alan gibi değişkenler kullanılır.



Doğrusal Regresyonun Matematiksel Temelleri

Bağımsız Değişken

Bağımsız değişken, regresyon denkleminde tahmin edilen değişkeni etkilediği düşünülen değişkendir. Örneğin, bir ev fiyatını tahmin ederken bağımsız değişkenler ev büyüklüğü, konum ve yaş olabilir.

Bağımlı Değişken

Bağımlı değişken, regresyon denklemi aracılığıyla tahmin edilen değişkendir. Örneğin, bir ev fiyatını tahmin ederken bağımlı değişken ev fiyatıdır.

Basit Doğrusal Regresyon Denklemleri

1 $Y = a + bX$

Bu denklemde, Y bağımlı değişken, X bağımsız değişken, a kesişim noktası (Y-ekseni ile kesişme noktası) ve b eğimdir.

2 Eğim

Eğim, bağımsız değişkenin her birim artması için bağımlı değişkenin ne kadar değiştığını temsil eder.

3 Kesişim Noktası

Kesişim noktası, bağımsız değişken sıfır olduğunda bağımlı değişkenin değerini gösterir.

Çoklu Doğrusal Regresyon Denklemleri

$$Y = a + b_1X_1 + b_2X_2 + \dots + b_nX_n$$

Bu denklemde, Y bağımlı değişken, X₁, X₂, ..., X_n bağımsız değişkenler, a kesişim noktası ve b₁, b₂, ..., b_n bağımsız değişkenlerin katsayılarıdır.

Katsayılar

Katsayılar, her bir bağımsız değişkenin bağımlı değişken üzerindeki etkisini ölçer.



Hata Terimleri ve Artık Analizleri



Hata Terimleri

Hata terimleri, modelin tahmin ettiği değerler ile gerçek değerler arasındaki farkı temsil eder.

Hata Terimlerinin Özellikleri:

Ortalama Sıfır Olmalıdır: Hataların toplamı sıfıra yakın olursa model doğru eğriye uyar.

Rastgele Dağılmalıdır: Hataların sistematik bir örüntü oluşturmaması gereklidir.

Normal Dağılım: Hata terimlerinin normal dağılım göstermesi beklenir.

Artık Analizi

Artık analizi, modelin tahmin ettiği değerler ile gerçek değerler arasındaki farkları inceleyerek modelin performansını değerlendirmek için kullanılır.

Amaç:

Hataların rastgele dağılıp dağılmadığını kontrol etmek.

Varyansın sabit olup olmadığını değerlendirmek (**homoskedastisite**).

Hata Terimlerinin Davranışı:

Eğer hatalar rastgele dağılmışsa model doğru çalışıyor demektir.

Eğer bir **örüntü** (örneğin bir eğri ya da artan/azalan trend) gözleniyorsa model doğrusal ilişkisi tam olarak yakalayamıyor olabilir.

Model Değerlendirme Ölçütleri

1 R-kare

Modelin varyasyonun ne kadarını açıkladığını ölçer, 0 ile 1 arasında değişir.

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2}$$

R-kare'nin Yorumu:

- R² = 1:** Model veriyi mükemmel bir şekilde açıklıyor.
- R² = 0:** Model verideki değişkenliği açıklayamıyor.
- 0 < R² < 1:** Model kısmen açıklıyor. R² değeri ne kadar yüksekse modelin performansı o kadar iyidir.

Örnek:

Bir modelin R² değeri **0.85** ise, bu model bağımlı değişkendeki değişkenliğin **%85'ini** açıklıyor demektir.

2 Ortalama Kare Hata (MSE)

Modelin tahmin hatalarının ortalama karesini ölçer.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2$$

Özellikleri:

- Hataları karelersek, pozitif ve negatif hatalar birbirini götürmez.
- Küçük MSE değeri, modelin tahminlerinin gerçek değerlere yakın olduğunu gösterir.

3 Kök Ortalama Kare Hata (RMSE)

Modelin tahmin hatalarının ortalama karesinin karekökünü ölçer.

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

Özellikleri:

- RMSE, hataların ortalamasını **aynı birim** üzerinden gösterir.
- Bu nedenle MSE'ye göre yorumu daha kolaydır.

Örnek:

Bir tahmin modelinde RMSE **5** ise, bu hata tahminlerin ortalama olarak gerçek değerlerden **5 birim** sapma gösterdiği anlamına gelir.

Özet Tablosu

Kavram	Tanım	Amaç
Hata Terimi	Hataları aynı birim üzerinden yorumlamak.	Model hatalarını ölçmek.
Artık Analizi	Hataların rastgele dağılımını ve varyansı incelemek.	Modelin uygunluğunu test etmek.
R-kare (R^2)	Modelin bağımlı değişkeni ne kadar açıkladığını gösterir.	Modelin açıklama gücünü değerlendirmek.
MSE (Ortalama Kare Hata)	Hataların karelerinin ortalaması.	Modelin doğruluğunu ölçmek.
RMSE (Kök Ortalama Kare Hata)	Hataların karekök ortalaması.	Hataları aynı birim üzerinden yorumlamak.



Doğrusal Regresyonun Varsayımları

1 **Doğrusal İlişki**

Bağımsız değişken ile bağımlı değişken arasında doğrusal bir ilişki olmalıdır.

2 **Bağımsızlık**

Hata terimleri bağımsız olmalıdır, yani bir hata terimi diğerlerinden etkilenmemelidir.

3 **Sabit Varyans**

Hata terimlerinin varyansı, bağımsız değişkenin değerinden bağımsız olarak sabit olmalıdır.

4 **Normal Dağılım**

Hata terimlerinin normal bir dağılıma sahip olması gereklidir.

Doğrusal Regresyon Uygulamaları



Finans

Hisselerin fiyatını tahmin etmek, riskleri değerlendirmek.



Sağlık

Hastalık riskini tahmin etmek, hastaların tedavilerine yanıtlarını tahmin etmek.



Pazarlama

Pazarlama kampanyalarının başarısını tahmin etmek, müşteri davranışını tahmin etmek.



Diğer

Emlak değerlerini tahmin etmek, satışları tahmin etmek.

Doğrusal Regresyonun Sınırları ve Gelecek

1

Sınırlamalar

Doğrusal regresyon, doğrusal olmayan ilişkilerle iyi başa çıkamaz.

2

Gelecek

Doğrusal regresyon, daha karmaşık modeller için temel bir araç olarak devam edecek.