Regresyon Giriş Özet

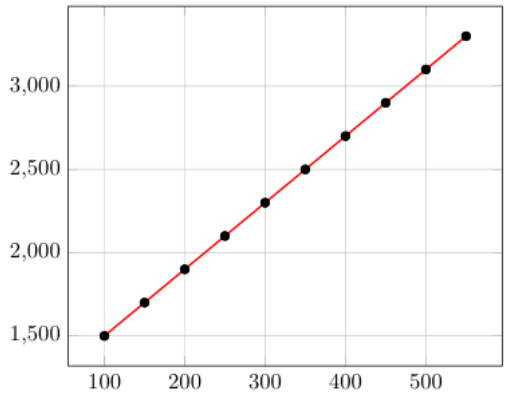
Gözetimli öğrenme, bilgisayara **doğru cevapları göstererek** bir görevi nasıl yapacağını öğretmektir; tıpkı bir öğretmenin öğrenciye ders anlatması gibi.
Gözetimsiz öğrenme ise, bilgisayarın kendi başına, verinin içindeki gizli desenleri ve ilişkileri keşfetmesini sağlamaktır; tıpkı bir kaşifin haritasız bir bölgeyi keşfetmesi gibi.

Basit Doğrusal Regresyon'a Giriş

Bu yöntem, biri bağımsız (girdi) ve diğeri bağımlı (çıktı) olmak üzere iki sürekli değişken arasındaki ilişkiyi modellendirir.

Reklam Harcaması ($)	Satış Geliri ($)
100	1500
150	1700
200	1900
250	2100
300	2300
350	2500
400	2700
450	2900
500	3100
550	3300

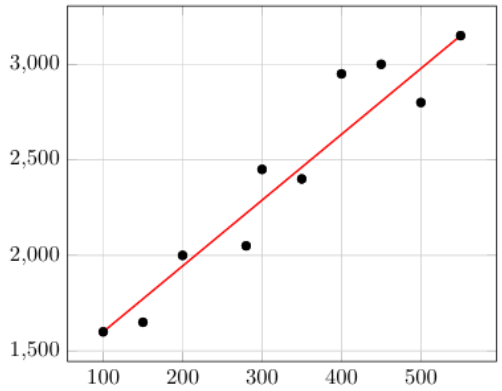
Peki, 475$'lık bir reklam harcaması yaptığımızda ne kadar gelir elde edebiliriz? İşte bunu tahmin etmek istiyoruz.



Şekil 1: Veri noktaları ve en iyi uyum doğrusu.

Burda doğrusal denklem grafiği görürüz. $y = \theta_0 + \theta_1 x$ şeklinde olur.

Ne yazık ki gerçek hayatta veriler mükemmel bir doğru üzerinde durmaz.



Şekil 2: Gürültü içeren veriler ve en iyi uyum doğrusu.

Gözlem: Veriler doğruya yakın ama birebir üzerinde değil. Bu yüzden hata kavramı doğar

Hataları Ölçmek

Bir tahminin ne kadar doğru olduğunu ölçmek için: $Hata = y - \widehat{y}$.

y = gerçek değer $\widehat{y}$ = tahmin edilen değer. Ve tüm hataların karesinin ortalamasını alırız.

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

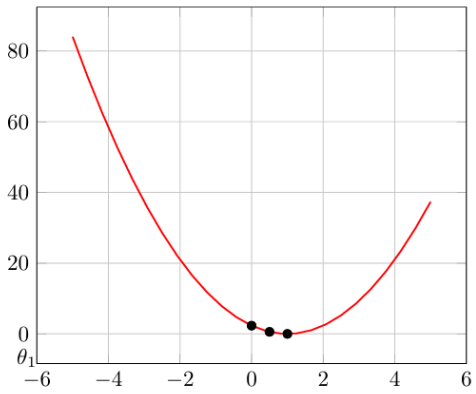
Bu değere **Mean Squared Error (MSE)** denir. Modelin amacı: MSE'yi mümkün olan en küçük değere indirmek.

Modeli İyileştirmek: Gradient Descent

Tahmin ile gerçek değer arasındaki farkı ölçerek bir hata (maliyet) hesaplıyoruz. Bu hatayı mümkün olduğunca küçük yapmak istiyoruz. İşte burada **Gradient Descent (Gradyan İnişi)** adlı algoritma devreye giriyor.

Şimdi bizim elimizde $y = \theta_0 + \theta_1 x$ denkleminden θ_0 ve θ_1 ayarlanabilir iki değer var. Eğer θ0 ve θ1 değerlerini doğru seçersek, tahminlerimiz daha iyi olur ve hata azalır.

EX: Mesela başlangıçta θ_0=100, θ_1=5 degeri var ve hata büyük. Bu değerleri küçük küçük değiştirerek hatayı azaltmaya çalışacağız.



Şekil 3: Maliyet fonksiyonu ve minimum hata noktası.

Grafikte, yatay eksende θ1 değerleri ve dikey eksende modelimizin hatası (MSE) var. Hedefimiz: θ1 = 1 değerine ulaşmak çünkü hata burada minimum (sıfır) oluyor.

Gradient Descent Güncelleme Formülü

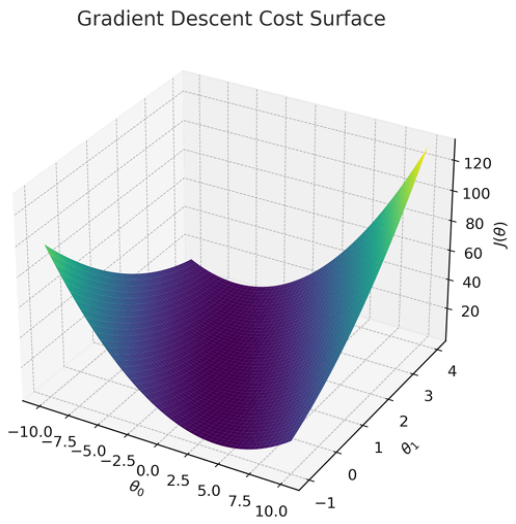
α = öğ$renme\ oran$, kesirli ifade ise eğim(ne tarafa inmeli)

$$\theta := \theta - \alpha \cdot \frac{\partial J(\theta)}{\partial \theta}$$

Çoklu Doğrusal Regresyon

Birden fazla girdimiz varsa kullandığımız modele Çoklu Doğrusal Regresyon (Multiple Linear Regression) denir.

Ya mesela ev fiyatı için lokasyon, oda sayısı, dayanıklılık gibi terimler fiyatını değiştirir değil mi. İşte bu durumda da→ $y = \theta_0 + \theta_1 x + \theta_2 x + \theta_3 x + \ldots$ şeklinde formulu vardır.



Şekil 4: Gradient Descent ile 3D maliyet yüzeyi.

Bu durumlar artık bir yüzeydir. Yüzeyin çukur noktası, **global minimum** noktasıdır.

Fakat basit de olsa çoklu da olsa prensip aynı kalır: **Maliyet fonksiyonunu minimize et!**

Doğrusal Regresyonda Model Başarısını Ölçme

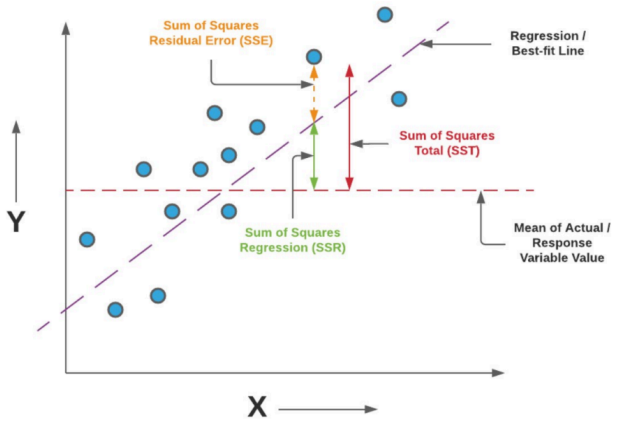
R-Kare ve Adjusted R-Kare

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

*__R-kare değeri__, modelin veriye ne kadar uyduğunu gösterir.

$SS_{res} = Gerçek\ değer\ ile\ tahmin\ arasndaki\ farkların\ karelerinin\ toplam\ (hata)$

$SS_{tot} = Gerçek\ değerlerin\ ortalamadan\ farklarnn\ karelerinin\ toplam\ (toplam\ varyans)$



R-kare, her yeni değer eklendiğinde artma eğilimindedir. Bu yanıltıcıdır. Eklenen değer sonuç ile alakasız ise R-kare artsa bile model aslında daha iyi olmamış olur. ***Adjusted R-kare***, bu alakasız değişkenleri cezalandırır ve gereksiz değişken eklenmesini engeller.

**n: veri #, p:bağımsız değişken sayısı.

$$R^2_{\text{adj}} = 1 - \left(1 - R^2\right) \cdot \frac{n-1}{n-p-1}$$

Model	R^2	Adjusted R^2
Model 1 (2 değişken)	0.88	0.85
Model 2 (3 değişken)	0.90	0.84
Model 3 (3 anlamlı değişken)	0.92	0.89

Yalnızca anlamlı değişkenlerin eklenmesiyle Adjusted R^2 artar.

Hata Metrikleri: MSE, MAE ve RMSE

- Bir regresyon modeli eğittikten sonra, ne kadar iyi tahmin yaptığını ölçmek için bazı metrikler kullanırız.

MSE;

Tahmin edilen değer ile gerçek değer arasındaki farkın karesinin ortalamasıdır.
*Hataların büyüklüğünü artırır bu sayede büyük hatalar büyük cezalandırılır.
*Türevlenebilir olduğundan **gradient descent** için uygundur.
**Aykırı değerler üzerinde büyük etkiye sahiptir.

MAE;

Tahmin ile gerçek değer arasındaki farkların mutlak değerlerinin ortalamasıdır.
*Hatanın birimi orijinal veriyle aynıdır.
*Aykırı değerlere karşı daha dayanıklıdır.
Her noktada türevlenebilir olmadığı için **gradient descent için zordur.

RMSE;

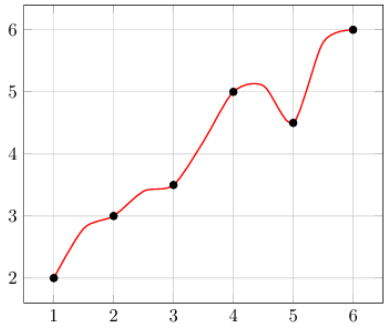
MSE nin karekökü dür. Hem karesini alır, hem de birim dönüşümü sağlar.

-Makine öğrenmesi modelleri eğitildikten sonra test verisinde nasıl davrandıklarına göre farklı sınıflandırma lara tabi tutulurlar.
-Overfitting (aşırı öğrenme)
-Underfitting (yetersiz öğrenme)
-Genelleştirilmiş model

Overfitting

Eğitim verisini neredeyse ezberlemiş ise ama yeni verilerde başarısız oluyor ise, bu durum overfitting olarak adlandırılır.

*Eğitim Doğruluğu→çok yüksek
*Test Doğruluğu→düşük
*Düşük bias, yüksek varyans

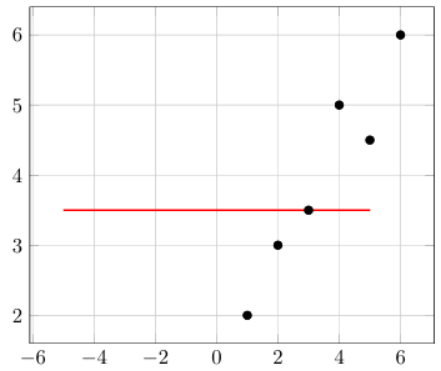


Şekil 1: Eğitim verisine aşırı uyan, ama genelleme yeteneği düşük model.

Underfitting

Ne eğitim verisini ne de test verisini iyi öğrenemez. Verideki ilişkileri yakalayamaz.

*Eğitim doğruluğu→düşük
*Test doğruluğu→düşük
*Yüksek bias, Düşük varyans

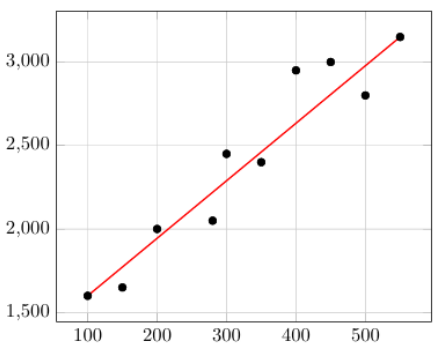


Şekil 2: Verideki yükselen ilişkiyi yakalayamayan basit model.

Genelleştirilmiş Model

Hem eğitim hem de test verisinde yüksek doğruluk veren modeller genelleştirilmiş kabul edilir.

*Eğitim doğruluğu→yüksek
*Test doğruluğu→yüksek
* Düşük bias, Düşük varyans



Şekil 3: Veriye iyi uyan ve genelleme yapabilen bir model.