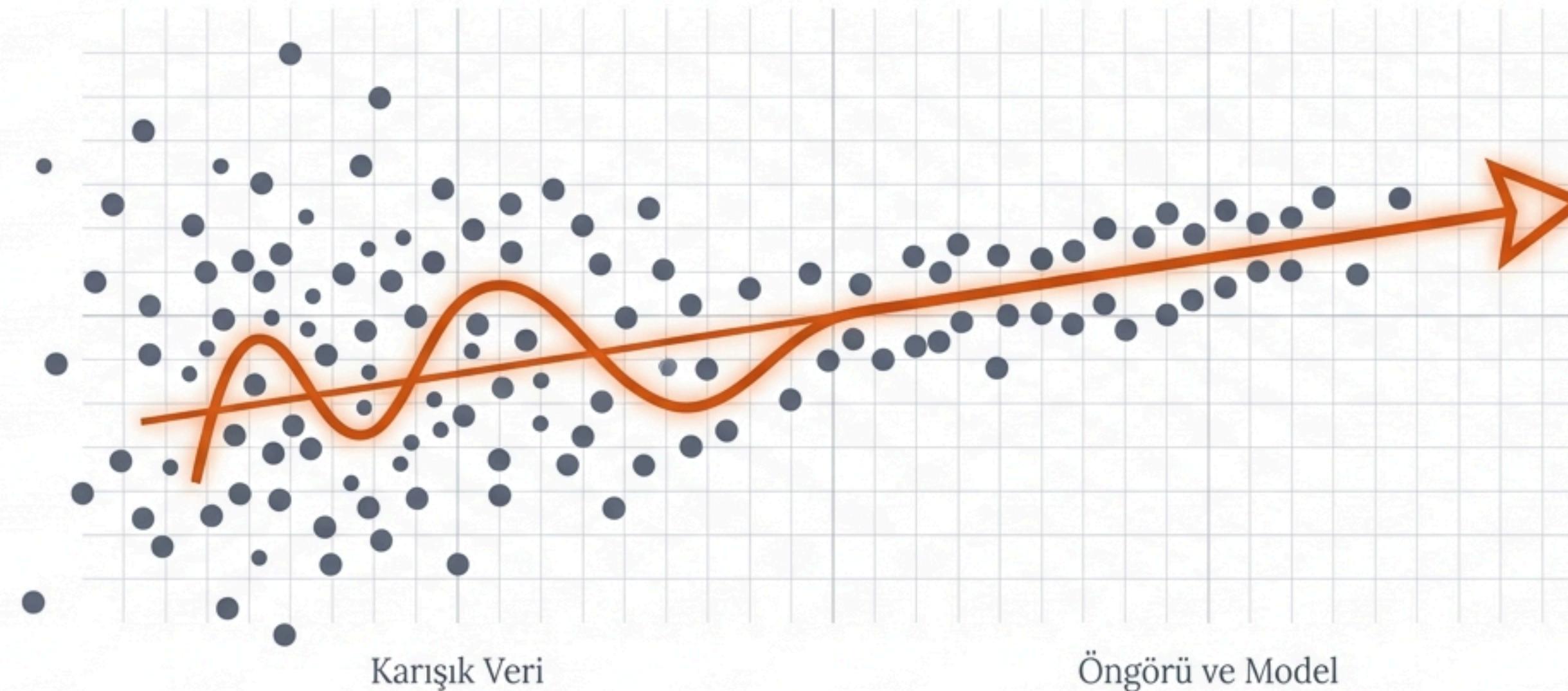


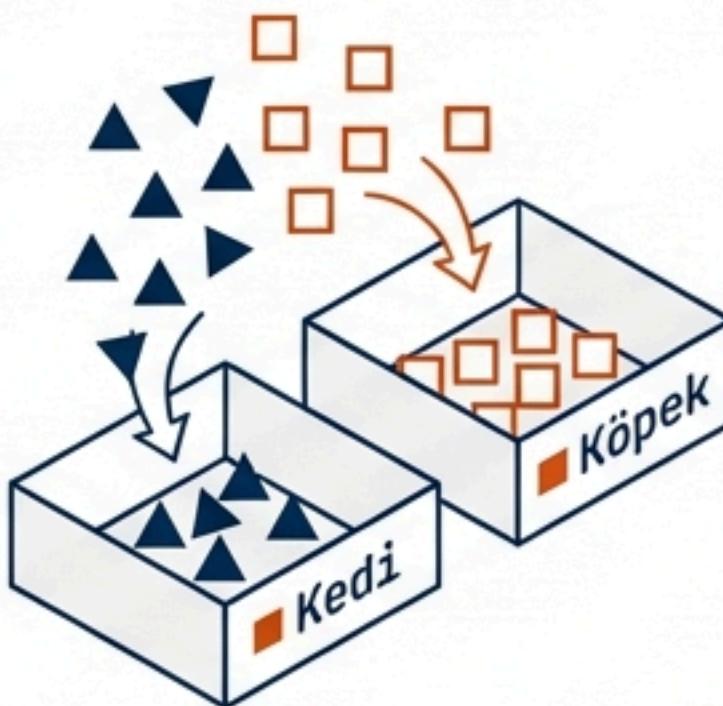
Regresyon Modelleriyle Geleceği Tahmin Etmek

Veriden İçgörüye: Doğrusal Modellerden Karmaşık Algoritmala Bir Yolculuk



Regresyon Nedir: Sınıflandırmadan Farkı

Sınıflandırma (Classification)



Verinin hangi kategoriye ait olduğunu bulur.
Örnek: Spam mail mi, değil mi?

Regresyon (Regression)



Bağımsız değişkenler ile hedef değişken arasındaki ilişkiyi modelleyerek sürekli sayısal bir değer tahmin eder.

Regresyon Örnekleri



Ev Fiyatı Tahmini

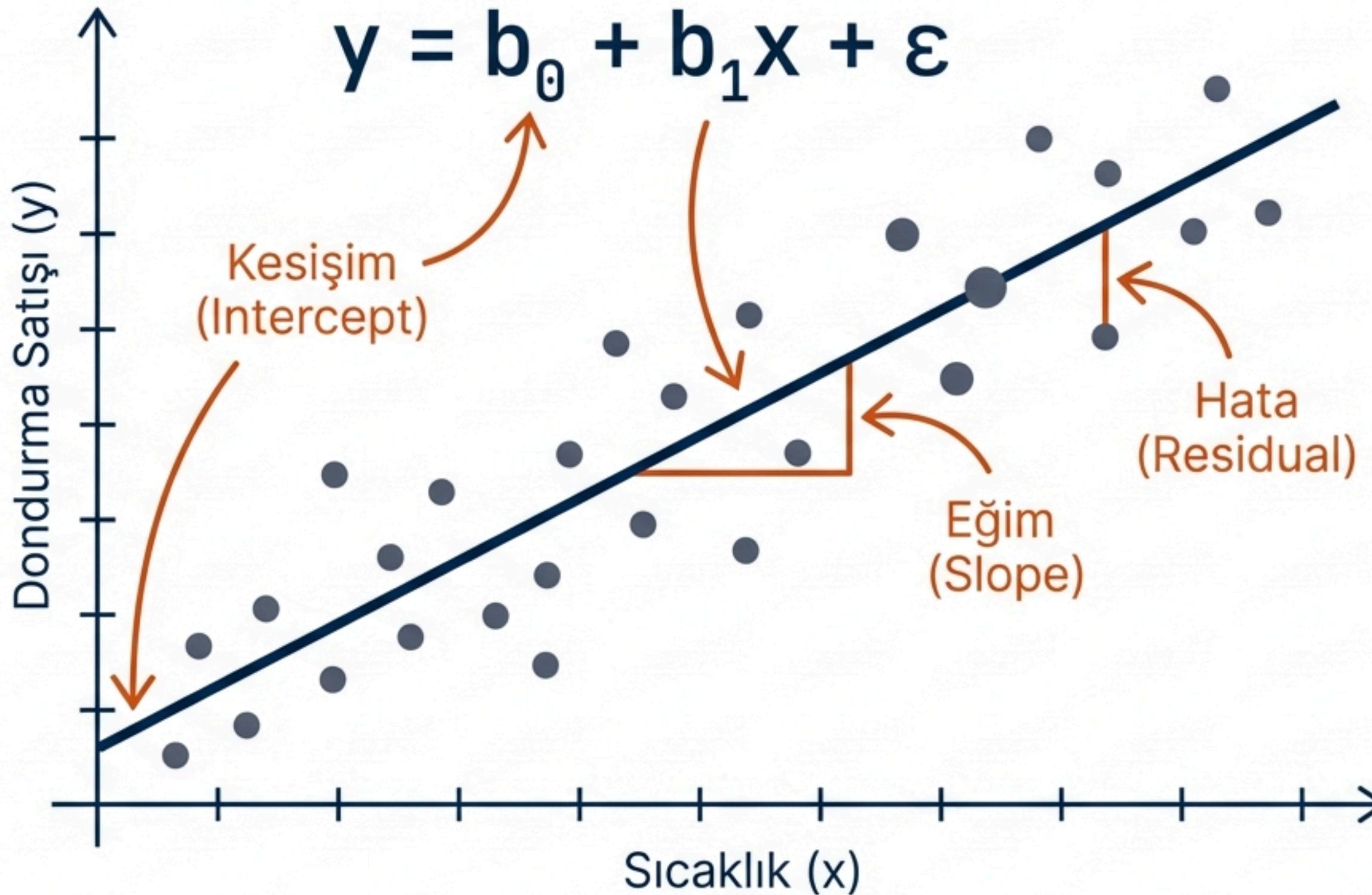


Sıcaklığa Göre Satış



Yakıt Tüketimi

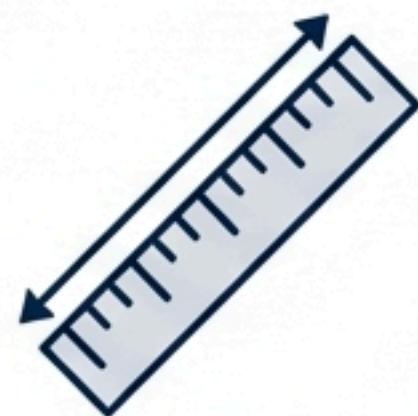
Temel Yapı Taşı: Basit Doğrusal Regresyon



OLS (Ordinary Least Squares):

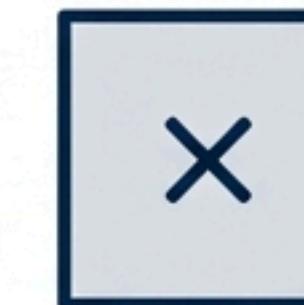
Amaç, tüm veri noktalarının çizgiye olan uzaklıklarının (hataların) kareleri toplamını toplamını minimize etmektir. En az hata, en iyi çizgiyi verir.

Başarıyı Nasıl Ölçeriz? Performans Metrikleri



MAE (Ortalama Mutlak Hata)

Hataların mutlak büyüklüğü. Yorumlaması en kolay metriktir. Hatalar birbirini götürmez, toplanır.



MSE (Ortalama Kare Hata)

Hataların karesini alır. Büyük hataları daha çok cezalandırır. Outlier'lara (aykırı değerlere) çok duyarlıdır.



RMSE (Kök Ortalama Kare Hata)

MSE'nin kareköküdür. Birimi orijinal veriyle (örneğin TL veya Derece) aynı hale getirir.



R² Skoru

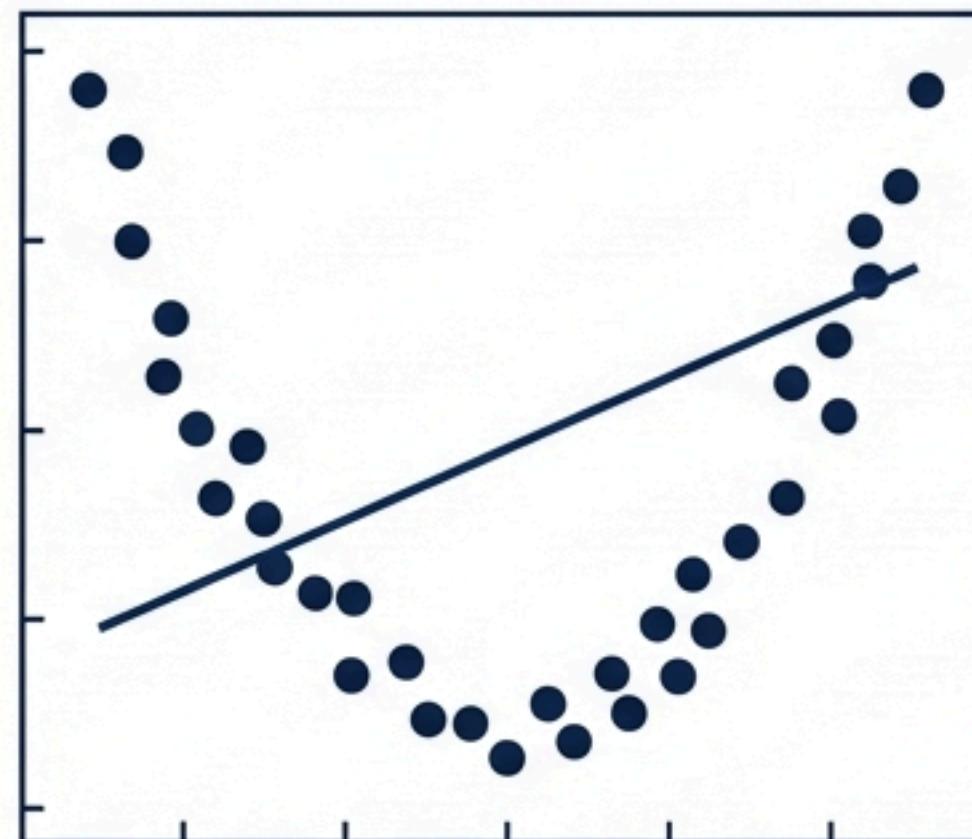
Modelin, bağımlı değişkendeki değişimi ne kadar iyi açıkladığını gösterir. (1'e ne kadar yakınsa o kadar iyi).

İpucu: Değişken sayısı arttıkça R² yapay olarak artabilir.
Bu durumda 'Adjusted R²' daha güvenilir bir metriktir.

İki Büyük Tuzak: Overfitting ve Underfitting

Bias (Yanlılık) ve Variance (Varyans) Dengesi

Underfitting (Yüksek Yanlılık)



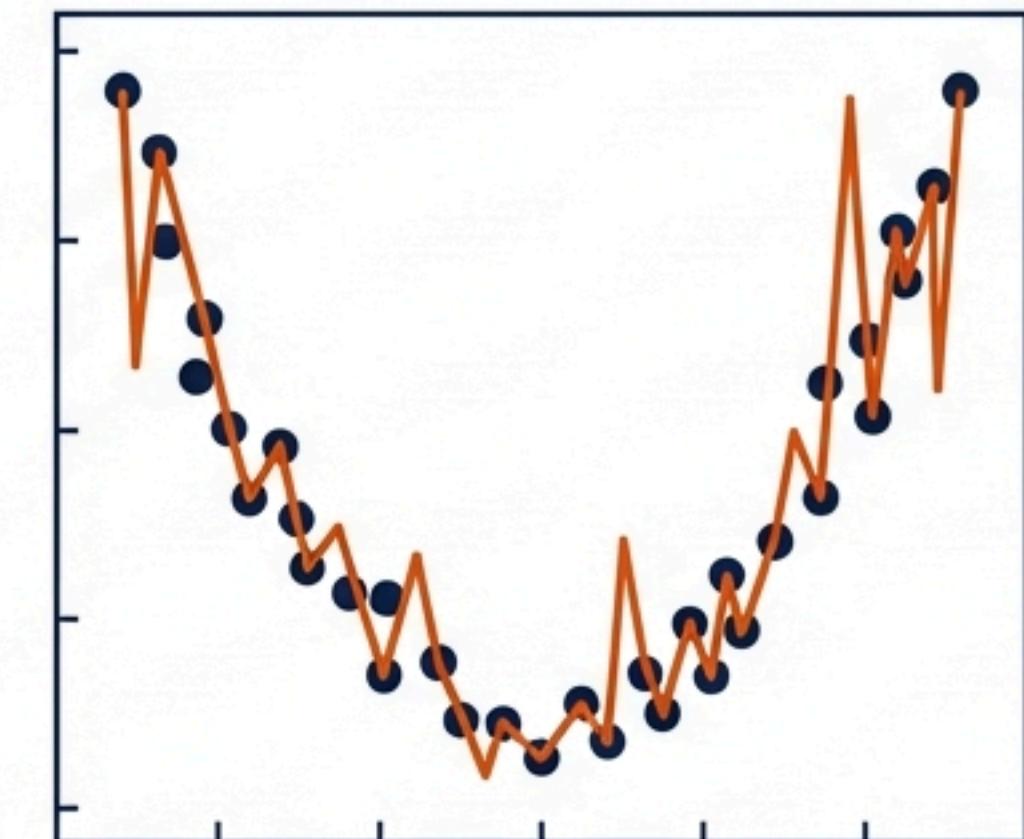
Model çok basit. Veriyi öğrenemiyor.

Optimal (Denge)



İdeal denge. Hem öğrenmiş hem genelleme yapabiliyor.

Overfitting (Yüksek Varyans)



Model veriyi ezberliyor. Gürültüyü bile modellemiş.

Veri Doğrusal Değilse? Polinomsal Regresyon

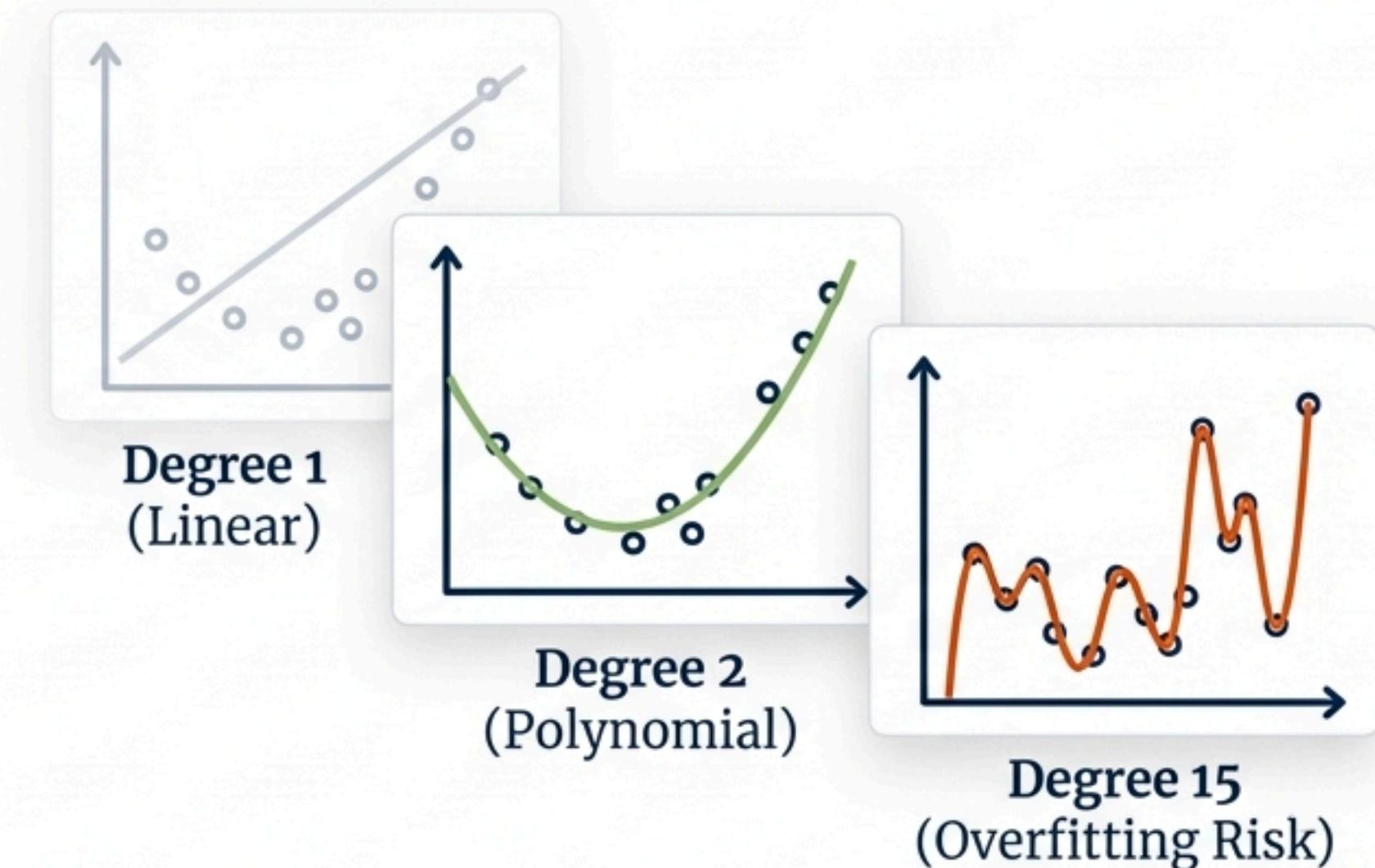
Sorun:

Veri seti parabolik veya kavisli bir yayılım gösterdiğinde (örneğin fırlatılan bir topun yörüngesi) düz bir çizgi yetersiz kalır (Underfitting).

Çözüm:

Denklemin derecesini (degree) artırmak.

$$y = b_0 + b_1x + b_2x^2 + b_3x^3 \dots$$



Dikkat: Derece çok artarsa model gürültüyü öğrenmeye başlar.

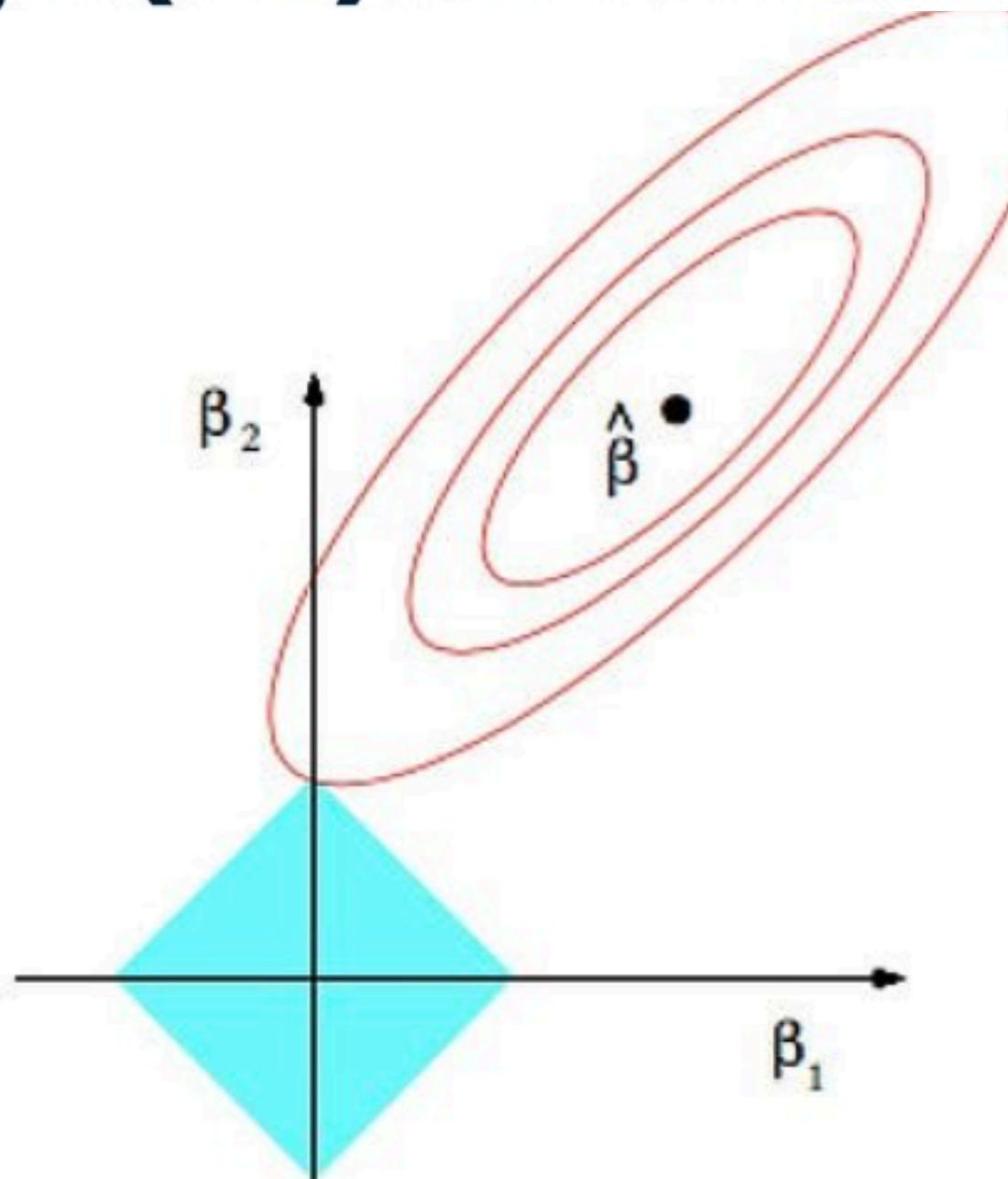
Karmaşıklığı Kontrol Altına Almak: Regülerizasyon



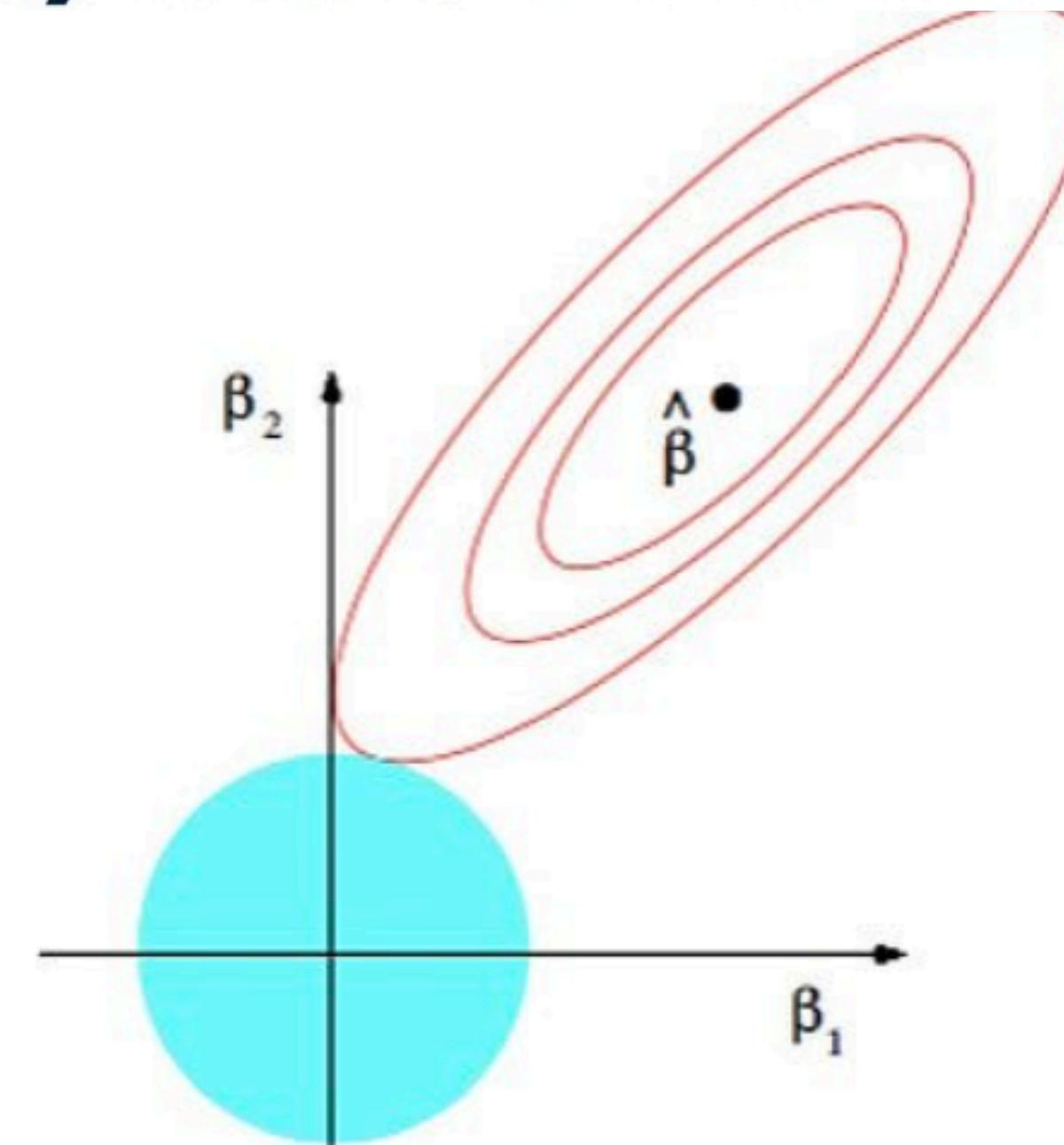
Nasıl Çalışır?

Hata fonksiyonuna ekstra bir maliyet eklenir. Bu, modelin katsayılarının aşırı büyümesini engeller. Sonuç olarak varyans düşer ve modelin genelleme yeteneği artar.

Ridge (L2) ve Lasso (L1) Farkı

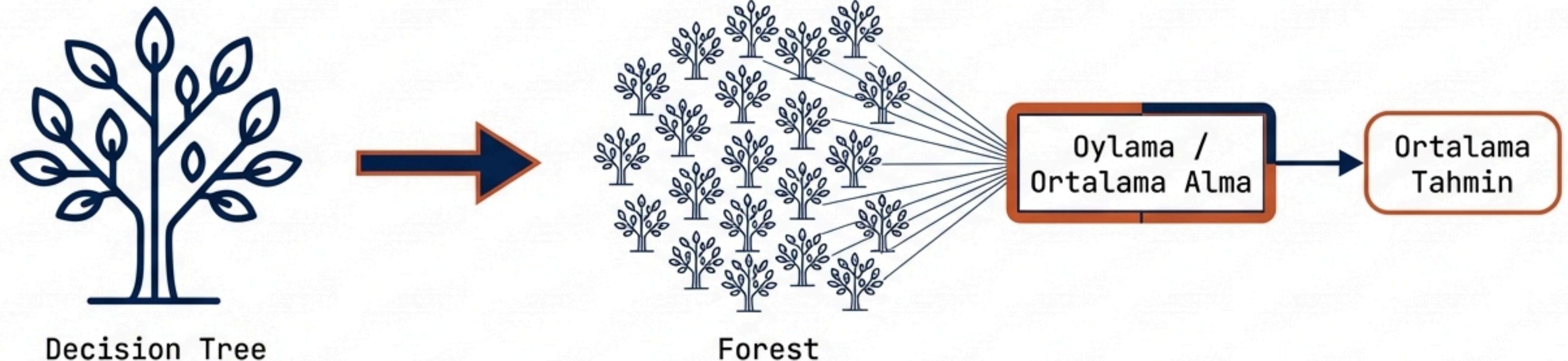


- Mutlak değer cezası ekler.
- Katsayıları **sıfıra indirebilir**.
- **Özellik Seçimi (Feature Selection)** yapar.
Gereksiz değişkenleri eler.



- Kare cezası ekler.
- Katsayıları sıfıra yaklaşır ama **sıfırlamaz**.
- **Multicollinearity** (ilişkili değişkenler) sorununda etkilidir.

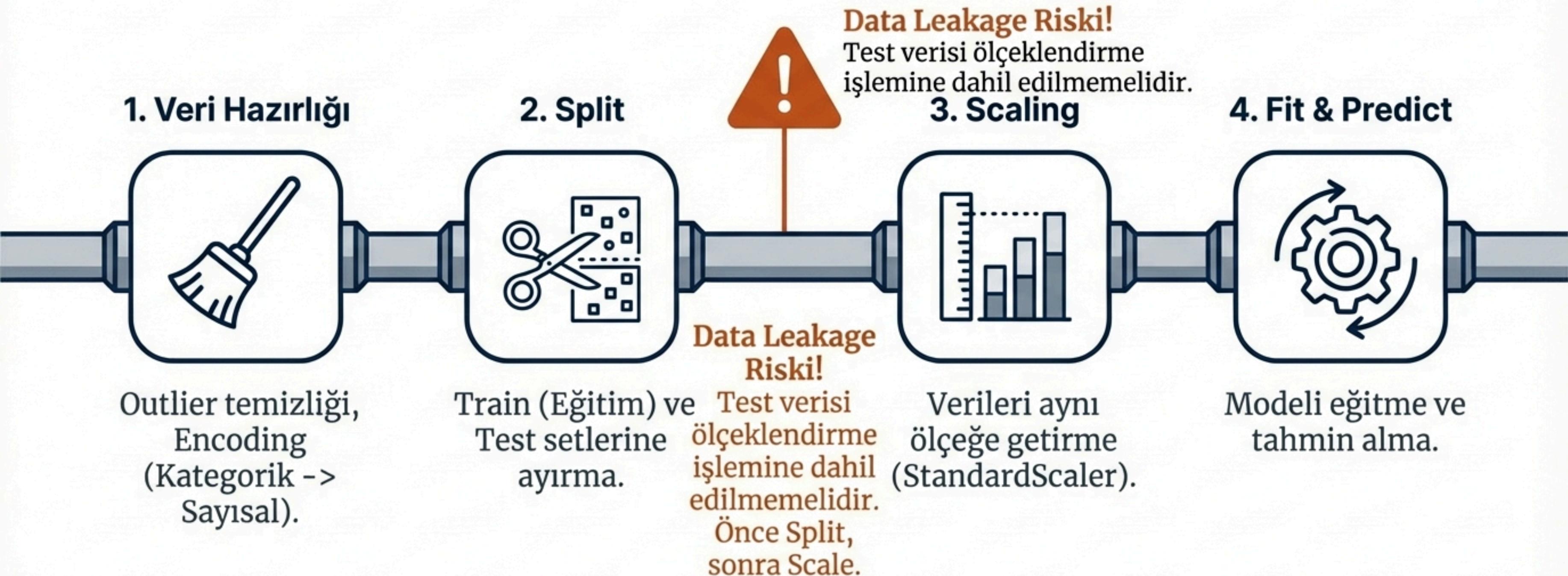
Doğrusallığın Ötesi: Random Forest ve Topluluk Öğrenmesi



Random Forest: Tek bir ağaç yerine yüzlerce ağacın ortak kararını (Ensemble Learning) kullanır.

- Doğrusal olmayan ve karmaşık ilişkileri yakalar.
- Kategorik verileri (Örn: 'Sigara içiyor mu?') daha iyi işler.
- Tek bir ağaca göre Overfitting riski daha düşüktür (Varyansı azaltır).
- Veriyi ölçeklendirmeye (Scaling) ihtiyaç duymaz.

Model Oluşturma İş Akışı (Pipeline)



Doğrusal Model Varsayımları

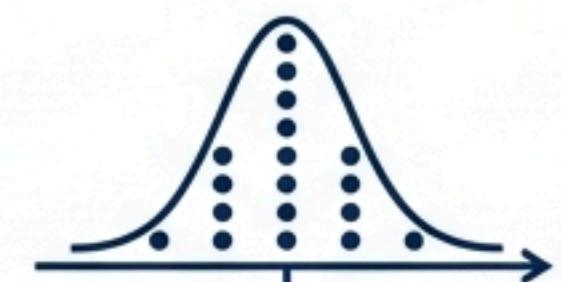
Modelin geçerli olması için gereken kurallar



Lineerlik: Bağımsız ve bağımlı değişken arasındaki ilişki doğrusal olmalıdır.



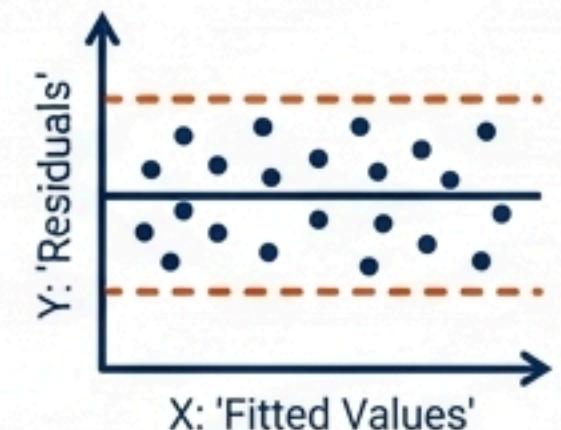
Normallik: Hatalar (residuals) normal dağılım göstermelidir.



Multicollinearity Yokluğu: Bağımsız değişkenler birbiriyle yüksek korelasyonlu olmamalıdır.

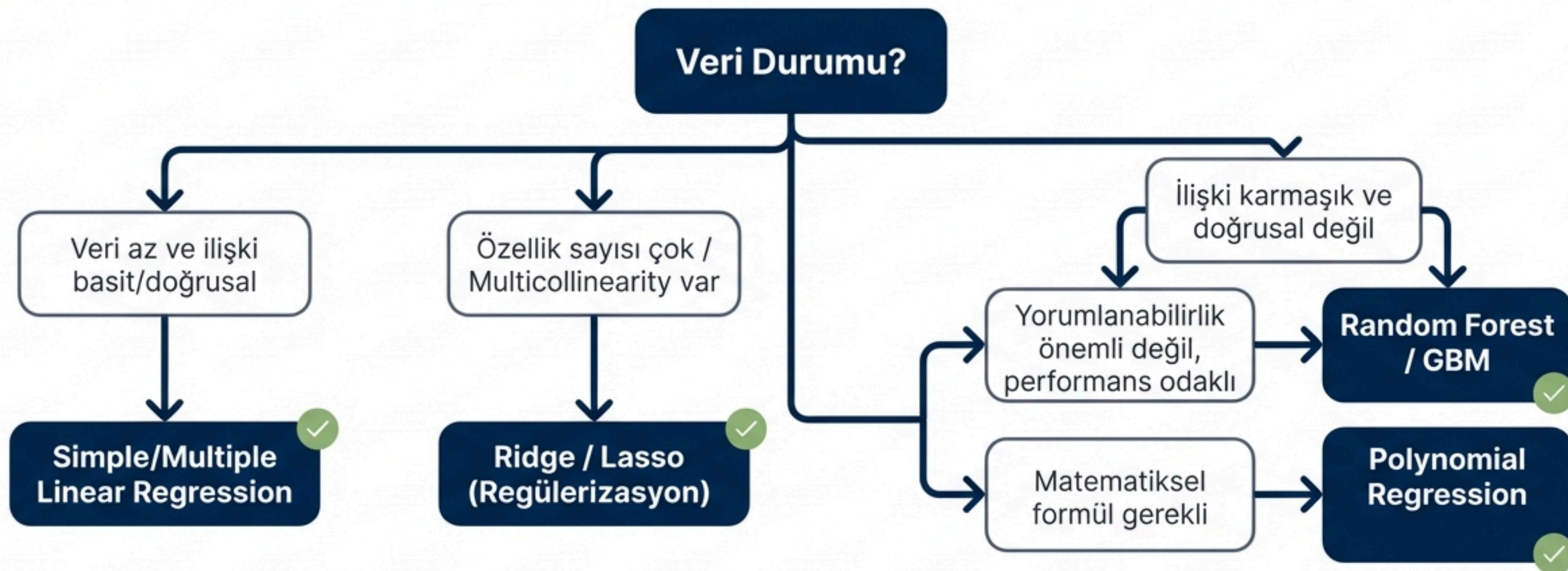


Homoscedasticity (Eş Varyanslık): Hataların varyansı veri seti boyunca sabit kalmalıdır.



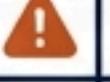
Otokorelasyon Yokluğu: Hatalar birbirinden bağımsız olmalıdır.

Karar Rehberi: Hangi Modeli Seçmeliyim?



! Not: Yorumlanabilirlik her zaman önemliyse Linear Modeller tercih edilir.

Modellerin Karşılaştırmalı Özeti

Model	Avantajlar	Dezavantajlar	Kullanım Alanı
Linear Regression	 Basit, hızlı, yorumlanabilir.	Karmaşık verilerde yetersiz (Underfit).	Basit trendler, satış tahminleri.
Polynomial Regression	 Kavisli ilişkileri modeller. 	Derece artarsa Overfit riski yüksek. 	Büyüme eğrileri, fiziksel olaylar.
Ridge / Lasso JetBrains Mono	 Overfitting'i engeller, genellenebilirlik. 	Parametre (Lambda) ayarı gerektirir. 	Çok değişkenli, gürültülü veriler.
Random Forest JetBrains Mono	 Yüksek doğruluk, outlier'lara dirençli. 	Yavaş, yorumlaması zor (Kara Kutu). 	Karmaşık, lineer olmayan problemler.

Kritik Çıkarımlar



Dengeyi Bulun

Model karmaşıklığı ile hata oranı arasında (Bias-Variance) optimal dengeyi arayın.



Veriyi Tanıyın

Outlier'lar var mı? İlişki doğrusal mı? Modeli veriye göre seçin, veriyi modele uydurmayın.



Körleşmeyin

Sadece R^2 'ye güvenmeyin; RMSE ve MAE gibi hata metriklerini mutlaka kontrol edin.



Hazırlık Her Şeydir

Başarılı bir modelin sırrı, doğru yapılan veri temizliği ve ölçeklendirmede saklıdır.