

Ridge, Lasso, Elastic Net Regresyon

-Doğrusal regresyon modelinde amacımız best fit line'i bulmaktır. Fakat bazen bu çizgi eğitim verisine çok iyi uyarsa overfitting olur ve yeni veri geldiğinde kötü sonuçlar verebilir.

Ridge Regresyon (L2 Regularization)

Doğrusal regresyon modeline L2 regularization ekleyen bir yöntemdir. Bu sayede modelin ağırlıkları (katsayıları) küçültülerek overfitting engellenmeye çalışılır.

Lineer Regresyon Maliyet Fonksiyonu:

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

Ridge Regresyon Maliyet Fonksiyonu:

$$J_{ridge}(\theta) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2 + \lambda \sum_{j=1}^n \theta_j^2$$

-Buradaki **λ değeri, regülarizasyon katsayısıdır.**
 λ büyüdükçe modelin ağırlıkları küçülür.
- θ ise best fit line 'deki denklemin katsayıları.

*Aykırı değerlere karşı dayanıklıdır.

*Özelliklerin çıktı üzerindeki etkisi azsa, bu etkileri bastırır.

*Çoklu doğrusal regresyonda korelasyon sorunlarını azaltır.

λ (lambda) ve Katsayılar arasındaki ilişki

$\lambda=0$ →Ridge, klasik doğrusal regresyon olur.

λ arttıkça, Ağırlıklar (θ) Küçülür: Düzenileştirme (ceza) arttıkça; model, katsayı değerlerinin mümkün olduğunca 0'a yakın olmasını ister. Bu, **modelin karmaşıklığını azaltır** ve aşırı öğrenmeyi (overfitting) engeller.

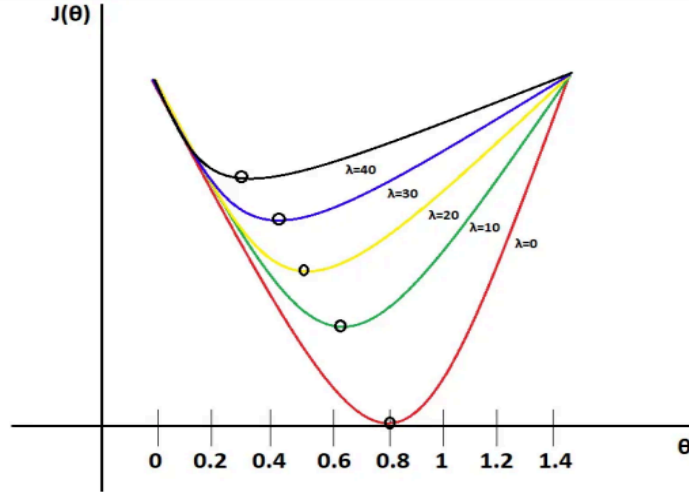
Katsayılar Nasıl Azalır?

Model, hem hatayı (MSE) hem de ceza terimini aynı anda en aza indirmeye çalışır.

Diyelim ki, modelin bir katsayısı (θ_j) çok büyük bir değer aldı (örneğin, 100). Bu, modelin tahmin hatasını azaltmaya yardımcı olabilir, ancak **ceza terimini ($\lambda \cdot 100^2$) çok büyük bir şekilde artırır.**

Model, toplam maliyeti (J_{ridge}) minimuma indirmek için, bu büyük katsayıyı küçültmeye başlar. Çünkü katsayıyı küçülttüğünde, tahmin hatası biraz artsa bile, **ceza terimindeki düşüş çok daha büyük olabilir** ve toplam maliyet fonksiyonu azalır.

- model MSE yi (ve ceza terimini) azaltmak için katsayıları düşürüyor.



Şekil 1: Lambda değeri arttıkça minimum noktası sola kayar, model daha küçük θ değerlerini tercih eder.

Lasso Regresyon (L1)

Hem aşırı öğrenmeyi (overfitting) azaltır hem de özellik seçimi (feature selection) sağlar. Bu yönüyle, özellikle çok sayıda değişkene (özellikle) sahip veri setlerinde oldukça etkilidir.

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m (\hat{y}_i - y_i)^2 + \lambda \sum_{j=1}^n |\theta_j|$$

L1 cezası bazı θ_j ağırlıklarının tam olarak sıfır olmasına neden olur;

*Eğer bir değişken hedef (bağımlı) değişkenle yeterince ilişkilendirilmemiş ise, θ_j sıfıra çekilir.

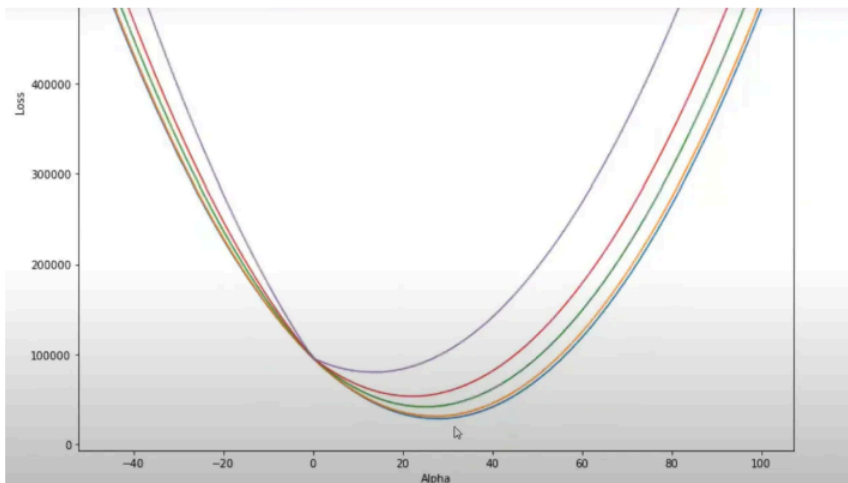
*Bu sayede önemsiz değişkenler modelden otomatik olarak çıkarılmış olur.

*Böylece model daha sade, daha açıklayıcı ve genellenebilir hale gelir.

-Çok sayıda özellik olduğu zaman kullanılabilir.

-Bazı özelliklerin anlamsız veya etkisiz olduğunu anlamlandırdığımızda

-Aşırı kullanmak underfitting'e yol açabilir.



Elastic Net

$$J(\theta) = \frac{1}{2m} \sum_{i=1}^m (\hat{y}_i - y_i)^2 + \lambda_1 \sum_{j=1}^n \theta_j^2 + \lambda_2 \sum_{j=1}^n |\theta_j|$$

Elastic Net, Ridge ve Lasso'nun avantajlarını birleştiren bir tekniktir. Hem aşırı öğrenmeyi azaltır hem de gereksiz değişkenleri ortadan kaldırır.

sonuç olarak;

Ridge, Lasso ve Elastic Net regresyonları, doğrusal modellerin daha kararlı ve anlamlı olmasını sağlar. Özellikle çok sayıda değişken içeren veri setlerinde Lasso ve Elastic Net'in değişken seçimi yetenekleri önemli avantaj sağlar. Modelin doğruluğunu artırmak için bu yöntemlerin doğru şekilde seçilmesi gerekir.