

Ridge, Lasso, Elastic Net Regresyon

Atıl Samancıoğlu

1 Giriş

Doğrusal regresyon modelinde, amacımız veriye en iyi uyan doğrusal bir çizgi (*best fit line*) bulmaktır. Ancak bazen bu çizgi eğitim verisine çok iyi uyarsa, yeni verilerle test edildiğinde kötü sonuçlar verebilir. Bu duruma **overfitting** (**aşırı öğrenme**) denir.

2 Overfitting Problemi

Eğer bir model eğitim verisine çok iyi uyum sağlıyorsa (örneğin doğruların tüm noktaların içinden geçtiği bir çizgi), ancak test verisinde büyük hatalar yapıyorsa, bu model genelleme yapamıyor demektir.

- Eğitim doğruluğu: Yüksek (hatta %100)
- Test doğruluğu: Düşük
- Bias: Düşük
- Varyans: Yüksek

3 Ridge Regresyon (L2 Regularizasyonu)

Aslında ceza verme yöntemiyle overfitting'e karşı önlem almaktır

Ridge regresyon, doğrusal regresyon modeline **L2 regularizasyon** ekleyen bir yöntemdir. Bu sayede modelin ağırlıkları (kat sayıları) küçültülerek overfitting engellenmeye çalışılır.

Lineer Regresyon Maliyet Fonksiyonu:

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

Overfitting'i engellemeye çalışırken katsayıları aşırı küçültürsek bu sefer de underfitting yapma ihtimalimiz doğuyor buna dikkat etmemiz gerekiyor

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 lambda ceza katsayısı olarak da düşünülebilir

Buradaki λ değeri, regularizasyon katsayısıdır. λ büyüdükçe modelin ağırlıkları küçülür. Lambda'nın çarptığı toplamın içinde ise slope'un (eğim) karesini görürüz.

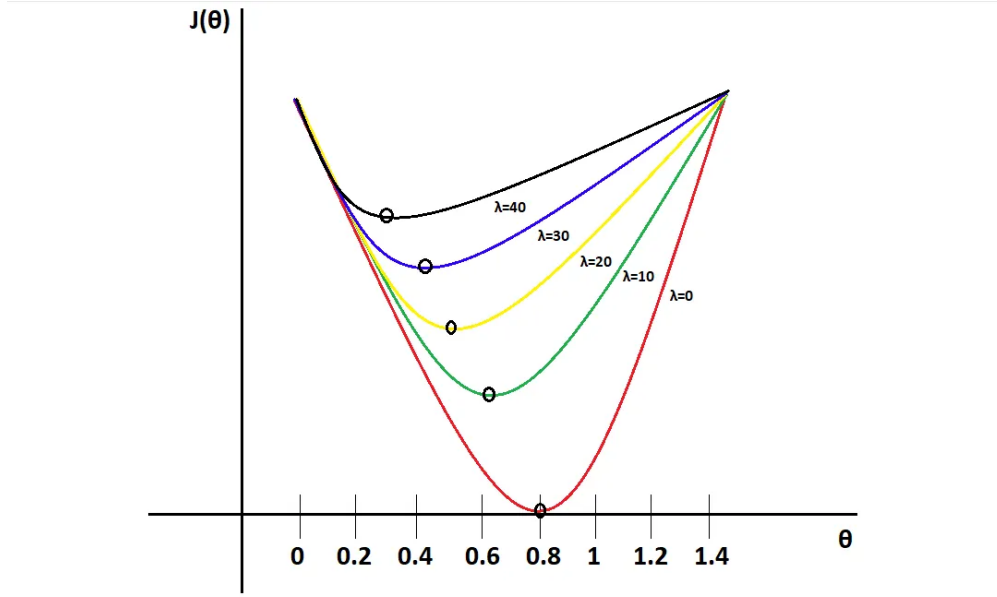
4 Neden Ridge Kullanılır?

- Aykırı değerlere karşı dayanıklıdır.
- Özelliklerin çıktı üzerindeki etkisi azsa, bu etkileri bastırır.
- Çoklu doğrusal regresyonda korelasyon sorunlarını azaltır.

5 Lambda ve Katsayılar Arasındaki İlişki

- $\lambda = 0$ iken Ridge, klasik doğrusal regresyon olur.
- λ arttıkça ağırlıklar (θ_j) küçülür.
- $\lambda \rightarrow \infty$ durumunda tüm ağırlıklar sıfıra yaklaşır ancak **tam sıfır olmaz**.

figure



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

6 Sonuç

Ridge regresyon:

- Aşırı öğrenmeyi azaltır.
- Özelliklerin katsayılarını küçültür, sıfırlamaz.
- L2 regularizasyonu ile model karmaşıklığını kontrol eder.
- λ hiperparametresi ile model ayarlanabilir.

7 Lasso Regresyon (L1)

Lasso (Least Absolute Shrinkage and Selection Operator) regresyon, hem **aşırı öğrenmeyi (overfitting)** azaltır hem de **özellik seçimi (feature selection)** sağlar. Bu yönüyle, özellikle çok sayıda değişkene (özellikle) sahip veri setlerinde oldukça etkilidir.

Bu da bir ceza yöntemi buradaki amaç daha çok giren feature'lar için çıktığı etkilemeyen değerleri bir nevi öğrenimden atması demek

Maliyet Fonksiyonu

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

Bu formülde, λ ceza (penalizasyon) katsayısıdır ve ikinci terim modelin ağırlıklarını (slope değerlerini) küçültmeye zorlar. Ridge regresyondan farkı, ağırlıkların karesi yerine mutlak değerlerinin kullanılmasıdır.

L1 Cezasının Etkisi: Özellik Seçimi

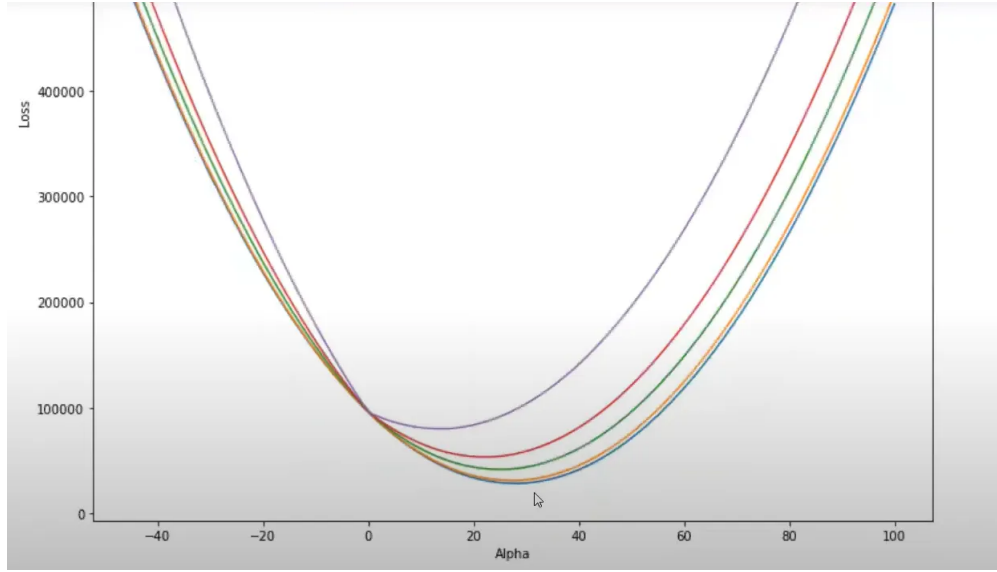
L1 cezası bazı θ_j ağırlıklarının **tam olarak sıfır** olmasına neden olur. Bu, şu anlama gelir:

- Eğer bir değişken hedef (bağımlı) değişkenle yeterince ilişkilendirilmemişse, θ_j sıfıra çekilir.
- Bu sayede önemsiz değişkenler modelden otomatik olarak çıkarılmış olur.
- Böylece model daha sade, daha açıklayıcı ve genellenebilir hale gelir.

Ne Zaman Kullanılır?

- Çok sayıda özelliğe (feature) sahip modellerde,
- Bazı değişkenlerin aslında anlamsız ya da etkisiz olabileceğinden şüphelenildiğinde,
- Modelin hem tahmin performansı hem de yorumlanabilirliği artırılmak istendiğinde kullanılır.

Not: Lasso regresyon çok fazla değişkeni sınırlayabilir. Bu durum bazı önemli ama zayıf etkili değişkenlerin de modelden atılmasına neden olabilir. Bu gibi durumlarda Elastic Net tercih edilebilir.



Şekil 2: Lasso regresyonda lambda arttıkça bazı ağırlıklar sıfıra indirgenebilir.

8 Elastic Net $L1+L2$

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

Maliyet Fonksiyonu

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

Özellikleri:

- λ_1 Ridge etkisi yaratır (L2)
- λ_2 Lasso etkisi yaratır (L1)
- Aşırı öğrenmeyi azaltırken gereksiz özellikleri sıfırlayabilir.

9 Karşılaştırma Tablosu

Özellik	Ridge	Lasso	Elastic Net
Regülerizasyon tipi	L2	L1	L1 + L2
Aşırı öğrenmeyi azaltma	Evet	Evet	Evet
Değişken seçimi	Hayır	Evet	Evet
Katsayıyı sıfıra indirme	Hayır	Evet	Evet

10 Sonuç

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.