

ELASTİCNET REGRESYON ALGORİTMASI

$$SSE_{Enet} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \boxed{\lambda_1} \sum_{j=1}^P \beta_j^2 + \boxed{\lambda_2} \sum_{j=1}^P |\beta_j|$$

The diagram illustrates the Elastic Net regression formula. The first term is the sum of squared residuals. The second term is the L2 penalty, represented by $\lambda_1 \sum \beta_j^2$, and the third term is the L1 penalty, represented by $\lambda_2 \sum |\beta_j|$. Arrows point from the text 'L2 ve L1 Ayar Parametleri' to the boxed parameters λ_1 and λ_2 . Another set of arrows points from the boxed parameters to the text 'Ceza Terimleri' (Penalty Terms).

Görsel 1.0

Elastic Net Nedir?

ElasticNet regresyonu, lineer regresyonun bir genişlemesi olarak kullanılan bir makine öğrenimi algoritmasıdır. ElasticNet regresyonu, 2005 yılında Hui Zou ve Trevor Hastie tarafından geliştirilmiştir.

Elastic Net, doğrusal regresyonun bir türüdür ve L1 (Lasso) ve L2 (Ridge) düzenlemelerini birleştirerek kullanır. Bu, ağırlıkları sıfıra çekme ve çoklu korelasyon özellikleri ele alabilme yeteneklerini birleştirir. İki farklı düzenleme terimini içerdiği için “elastic” adını alır.

Temelde, Elastic Net bir bağımlı değişkenin bir dizi bağımsız değişken tarafından doğrusal bir kombinasyonu ile tahmin edilmeye çalışıldığı bir regresyon modelidir. Ancak, geleneksel doğrusal regresyona ek olarak Elastic Net şu iki düzenleme terimini içerir:

L1 Düzenleme (Lasso): Bu terim, ağırlıkların toplamının bir kısmını sıfıra çeker. L1 düzenleme, özellik seçimini teşvik eder ve bazı ağırlıkları tamamen sıfıra indirebilir. Bu, gereksiz veya zayıf özelliklerin modelden çıkarılmasına yardımcı olur.

L2 Düzenleme (Ridge): Bu terim, ağırlıkların karesinin toplamının bir kısmını sıfıra çeker. L2 düzenleme, çoklu doğrusal korelasyonlu özellikleri ele alabilen bir düzenleme türüdür. Yüksek korelasyon durumlarında L2 düzenleme ağırlıkları dengeler ve ağırlıkları sıfıra çekmek yerine birbirine yakın tutmaya çalışır.

Elastic Net' in amacı, L1 ve L2 düzenleme terimlerini birleştirerek her iki yöntemin avantajlarını bir araya getirmektir. L1 düzenleme, seyrek modeller elde etme yeteneği ile öne çıkarken, L2 düzenleme çoklu korelasyonlu özelliklere karşı daha dirençlidir. Elastic Net, bu iki düzenleme terimlerini bir dengeleyici olarak kullanır ve regresyon katsayılarını optimize ederek esnek bir çözüm sunar.

Elastic Net'in temel formülü Görsel 1.0' da verilmiştir. Formül hakkında:

λ_1 : L2 düzenleme parametresi (Ridge regresyonundaki alpha değeri)

λ_2 : L1 düzenleme parametresi (Lasso regresyonundaki alpha değeri)

n: Değişken sayısı

L1 ayar parametresi eğer 0 olursa L2 cezalandırılması, 1 olduğunda ise L1 cezalandırılması uygulanır. Dolayısıyla 0–1 arasında değişimi aslında cezalandırma metodlarının göreceli etkilerini ifade eder. Eğer ikisinin de dengeli bir şekilde etki etmesini istiyorsak 0.5 değerinde bırakmamız doğru olacaktır.

$$\sum_{j=1}^P \beta_j^2$$

Ridge regresyonu tarafından uygulanan katsayıların karelerinin toplamı

$$\sum_{j=1}^P |\beta_j|$$

Lasso regresyonu tarafından uygulanan katsayıların mutlak değerlerinin toplamı

P: Özellik ya da değişken sayısı

Elastic Net' in Avantajları

Elastic Net, L1 ve L2 regresyon yöntemlerinin bir kombinasyonu olduğu için bu iki yöntemin avantajlarını birleştirir. İşte Elastic Net' in avantajları:

Seyrek Model Üretme Yeteneği:

L1 düzenleme (Lasso), ağırlıkların birçoğunu sıfıra çekerek seyrek bir model elde etme yeteneğine sahiptir. Bu, gereksiz veya zayıf özellikleri modelden çıkarmaya ve özellik seçimini iyileştirmeye yardımcı olabilir. Elastic Net, bu avantajı korur.

Çoklu Korelasyon ile Başa Çıkma:

L2 düzenleme (Ridge), çoklu korelasyonlu özelliklerle başa çıkma yeteneğine sahiptir. Elastic Net, bu avantajı koruyarak çoklu korelasyonlu özellikleri kontrol etmede etkilidir.

Esnek Parametre Ayarı:

Elastic Net, iki hiper parametre olan α (alpha) ve ρ (rho) kullanarak L1 ve L2 düzenleme terimleri arasındaki dengeyi ayarlamak için esnek bir yol sağlar.

Yüksek Boyutlu Verilerle Başa Çıkma:

Elastic Net, yüksek boyutlu verilere sahip veri setleri için uygundur. Overfitting'i önler ve gözlemlerden daha fazla özellik bulunan durumlarla başa çıkar.

Seyrek Veri İçin Uygunluk:

Elastic Net, seyrek verilerle (birçok girişin sıfır olduğu veri setleri) uğraşırken, özellikle uygun ön işleme teknikleri ile birleştirildiğinde etkili olabilir.

Otomatik Özellik Seçimi:

Elastic Net, otomatik olarak özellik seçimi yapabilir. Bu, yalnızca bir alt kümenin tahmin görevi için ilgili olduğu veri setleriyle çalışırken özellikle kullanışlıdır.

Elastic Net' in Dezavantajları

Elastic Net algoritması, birçok durumda etkili olabilir, ancak bazı durumlar ve koşullar altında dezavantajlar yaşanabilir. İşte Elastic Net'in potansiyel dezavantajları:

Model Karmaşıklığı:

Elastic Net, L1 ve L2 düzenleme terimlerini birleştirir, bu nedenle modelin matematiksel yapısı karmaşıklığı artırabilir. Bu durum, modelin anlaşılabilirliğini azaltabilir.

Veri Boyutlarına Hassasiyet:

Küçük veri setlerinde, özellikle yüksek boyutlu özellik uzaylarında, Elastic Net'in performansı daha hassas hale gelebilir. Bu durum, modelin aşırı uyum veya aşırı uygunluk sorunlarına yol açabilir.

Tekrarlayan Özellikler:

Elastic Net, birbirleriyle güçlü bir şekilde korele olan özellikleri gruplamak için tasarlanmış olsa da, aynı zamanda bir özellik grubundaki bazı özellikleri seçme eğilimindedir. Bu durum, modelin tekrar eden özelliklere duyarlı olmasına neden olabilir.

Elastic Net'in bu dezavantajlarına rağmen, birçok durumda etkili bir regresyon yöntemi olarak kabul edilir ve özellikle yüksek boyutlu ve korele veri setlerinde kullanışlı olabilir. Değerlendirme yapmadan önce, özellikle veri setinin özelliklerini ve problemi anlamak önemlidir.

ElasticNet Regresyon Algoritması Örnek

SORU ÇÖZÜMÜ

X	Y
1	1
2	1
2	2
3	2

$$Y = \beta_1 \cdot X + \beta_0$$

$$p = 0,5 \quad \lambda = 0,3$$

$$\alpha = 0,2 \text{ (Kademeli iniş için kullanılacak)}$$

SORU: Yukarıdaki veriseti için ElasticNet Algoritmasını kullanarak bir model tasarlayınız. (Bir çevrim için yapınız.)

Çözüm:

$$\text{ElasticNet MSE} = \text{MSE}(y, \hat{y}) + \lambda \cdot p \sum_{j=1}^m |\beta_j| + \lambda \cdot \frac{1-p}{2} \sum_{j=1}^m \beta_j^2$$

$$(L) \text{ Loss} = \frac{1}{2n} \sum_{i=1}^n (\beta_1 \cdot X_i + \beta_0 - y_i)^2 + \lambda p (|\beta_0| + |\beta_1|) + \lambda \cdot \frac{1-p}{2} (\beta_0^2 + \beta_1^2)$$

$$\frac{\partial L}{\partial \beta_0} = \frac{1}{4} \sum_{i=1}^4 (\beta_1 \cdot X_i + \beta_0 - y_i) \pm \lambda \cdot p + (1-p) \cdot \beta_0$$

$$\frac{\partial L}{\partial \beta_1} = \frac{1}{4} \sum_{i=1}^4 (\beta_1 \cdot X_i + \beta_0 - y_i) \cdot X_i \pm \lambda \cdot p + |1-p| \cdot \beta_1$$

$$\left\{ \begin{array}{l} \beta_0^{\text{yeni}} = \beta_0^{\text{eski}} - \alpha \cdot \frac{\partial L}{\partial \beta_0} \\ \beta_1^{\text{yeni}} = \beta_1^{\text{eski}} - \alpha \cdot \frac{\partial L}{\partial \beta_1} \end{array} \right\} \text{--- GRADIENT DESCENT ---}$$

$$\beta_0 = 1 \quad \text{ve} \quad \beta_1 = 1$$

$$\begin{aligned} \beta_0^{\text{yeni}} = 1 - 0,2 \left[\frac{1}{4} [(1+1-1) + (2+1-1) + (2+1-2) + (3+1-2)] + 0,15 + 0,5 \right] &= 0,66 \end{aligned}$$

$$\beta_1^{\text{yeni}} = 1 - 0,2 \left[\frac{1}{4} \left[(1+1-1) \cdot 1 + (2+1-1) \cdot 2 + (2+1-2) \cdot 2 + (3+1-2) \cdot 3 \right] + 0,15 + 0,5 \right] = 0,31$$

İlk Durum

$$\beta_1 = 1 \quad \beta_0 = 1$$

$$\hat{y}_1 = 1+1 = 2 \quad ; \quad y_1 = 1$$

$$\hat{y}_2 = 2+1 = 3 \quad ; \quad y_2 = 1$$

$$\hat{y}_3 = 2+1 = 3 \quad ; \quad y_3 = 2$$

$$\hat{y}_4 = 3+1 = 4 \quad ; \quad y_4 = 2$$

$$\boxed{SSE = 1 + 4 + 1 + 4 = 10}$$

İkinci Durum

$$\beta_1 = 0,31 \quad \beta_0 = 0,66$$

$$\hat{y}_1 = 0,31 + 0,66 = 0,97 \quad ; \quad y_1 = 1$$

$$\hat{y}_2 = 0,62 + 0,66 = 1,28 \quad ; \quad y_2 = 1$$

$$\hat{y}_3 = 0,62 + 0,66 = 1,28 \quad ; \quad y_3 = 2$$

$$\hat{y}_4 = 0,93 + 0,66 = 1,59 \quad ; \quad y_4 = 2$$

$$\boxed{SSE = (0,03)^2 + (0,28)^2 + (0,72)^2 + (0,41)^2 = 0,7658}$$

$$10 > 0,7658$$

Daha güzel bir model elde ettik ve Overfit durumundan kurtulmuş olduk.

SORU ÇÖZÜMÜNDE DİKKAT EDİLECEK NOKTALAR (AÇIKLAMALAR)

$$SSE_{Enet} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 + \lambda_1 \sum_{j=1}^k \beta_j^2 + \lambda_2 \sum_{j=1}^k |\beta_j|$$

Genel Formül (Hata Fonksiyonu)

Daha kullanışlı olması için (Soru çözümünde) yukarıdaki hata fonksiyonunda birkaç düzenleme yapabiliriz.

$$\text{ElasticNet Loss} = \frac{1}{2n} \sum_{i=1}^n (\beta_1 \cdot x_i + \beta_0 - y_i)^2 + \lambda \cdot p (|\beta_0| + |\beta_1|) + \lambda \cdot \frac{1-p}{2} (\beta_0^2 + \beta_1^2)$$

* λ_1 ve λ_2 kullanmak yerine tek bir λ kullanabiliriz. ($\lambda_1 = \lambda_2 = \lambda$)

* L_1 (Lasso) ve L_2 (Ridge) oranı genellikle "p" sembolü ile temsil edilir. Bu değer, modelin L_1 (Lasso) düzenlemesi ile L_2 (Ridge) düzenlemesi arasındaki dengeyi kontrol eder. Birine "p", diğerine "1-p" diyebiliriz.

Çünkü toplamları 1 olmak zorundadır.

* $\lambda \cdot \frac{1-p}{2} (\beta_0^2 + \beta_1^2)$ burada " $\frac{1-p}{2}$ " yazmamızın nedeni türev aldıktan

sonra daha kolay sadeleştirme yapabilmek içindir.

VERİ KÜMESİ HAKKINDA BİLGİ

Bu veri seti, Portekiz 'Vinho Verde' şarabının kırmızı varyantlarıyla ilgilidir. Veri seti, şarapta bulunan çeşitli kimyasalların miktarını ve bunların şaraba olan etkisini açıklar. Veri setleri sınıflandırma veya regresyon görevleri olarak görülebilir. Sınıflar sıralıdır ve dengesizdir (örneğin, normal şaraplar mükemmel veya zayıf olanlardan çok daha fazladır).

Veri çerçevesi şu sütunları içeriyor:

Input variables (based on physicochemical tests):

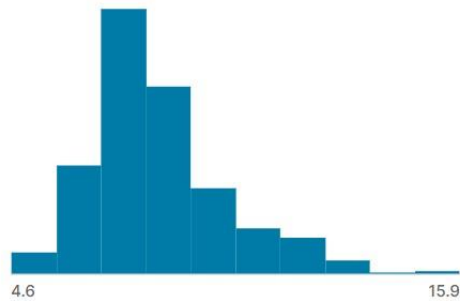
- 1 - fixed acidity(Sabit asitlik)
- 2 - volatile acidity(Uçucu asitlik)
- 3 - citric acid(Sitrik asit)
- 4 - residual sugar(Kalan şeker)
- 5 - chlorides(Klorürler)
- 6 - free sulfur dioxide(Serbest sülfür dioksit)
- 7 - total sulfur dioxide(Toplam sülfür dioksit)
- 8 - density(Yoğunluk)
- 9 - Ph
- 10 - sulphates(Sülfatlar)
- 11 - alcohol(Alkol)

Output variable (based on sensory data):

- 12 - quality {score between 0 and 10} (Kalite {score 0 ile 10 arasında})

fixed acidity

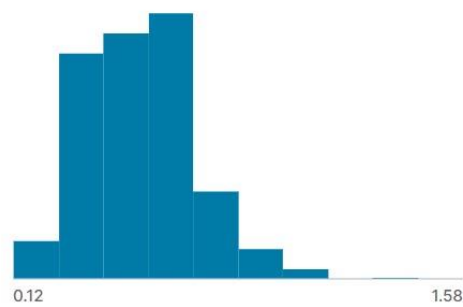
Fixed acidity value



Valid	1143	100%
Mismatched	0	0%
Missing	0	0%
Mean	8.31	
Std. Deviation	1.75	
Quantiles	4.6	Min
	7.1	25%
	7.9	50%
	9.1	75%
	15.9	Max

volatile acidity

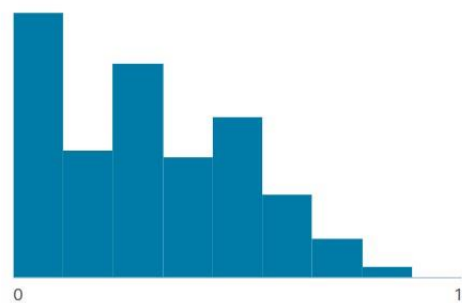
Volatile acidity value



Valid	1143	100%
Mismatched	0	0%
Missing	0	0%
Mean	0.53	
Std. Deviation	0.18	
Quantiles	0.12	Min
	0.39	25%
	0.52	50%
	0.64	75%
	1.58	Max

citric acid

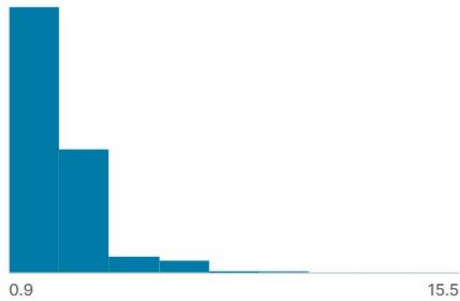
Citric acid value



Valid	1143	100%
Mismatched	0	0%
Missing	0	0%
Mean	0.27	
Std. Deviation	0.2	
Quantiles	0	Min
	0.09	25%
	0.25	50%
	0.42	75%
	1	Max

residual sugar

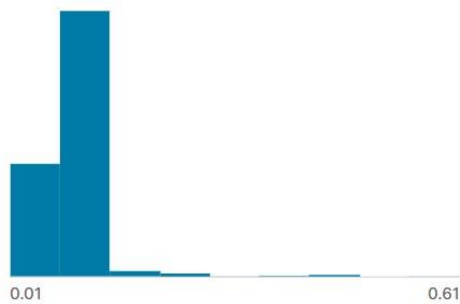
Residual sugar value



Valid	1143	100%
Mismatched	0	0%
Missing	0	0%
Mean	2.53	
Std. Deviation	1.36	
Quantiles		
	0.9	Min
	1.9	25%
	2.2	50%
	2.6	75%
	15.5	Max

chlorides

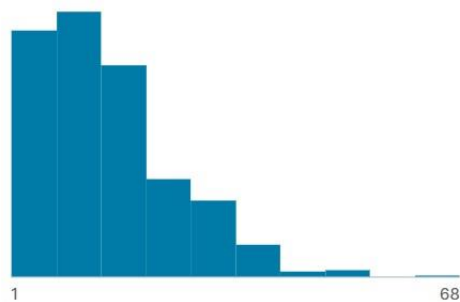
Chlorides value



Valid	1143	100%
Mismatched	0	0%
Missing	0	0%
Mean	0.09	
Std. Deviation	0.05	
Quantiles		
	0.01	Min
	0.07	25%
	0.08	50%
	0.09	75%
	0.61	Max

free sulfur dioxide

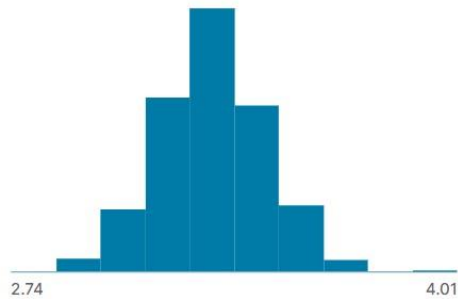
Free sulfur dioxide value



Valid	1143	100%
Mismatched	0	0%
Missing	0	0%
Mean	15.6	
Std. Deviation	10.2	
Quantiles		
	1	Min
	7	25%
	13	50%
	21	75%
	68	Max

pH

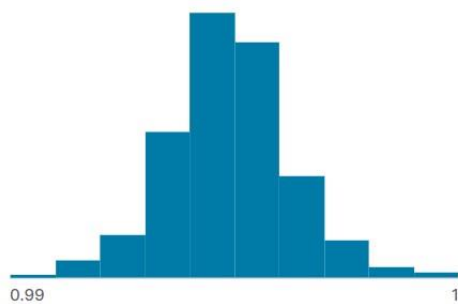
pH value



Valid	1143	100%
Mismatched	0	0%
Missing	0	0%
Mean	3.31	
Std. Deviation	0.16	
Quantiles	2.74	Min
	3.2	25%
	3.31	50%
	3.4	75%
	4.01	Max

density

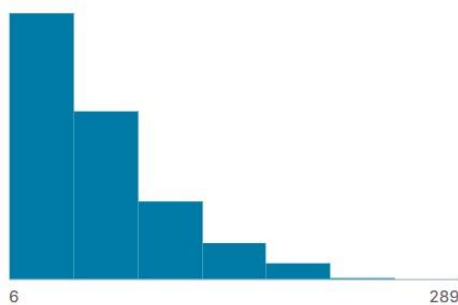
Density value



Valid	1143	100%
Mismatched	0	0%
Missing	0	0%
Mean	1	
Std. Deviation	0	
Quantiles	0.99	Min
	1	25%
	1	50%
	1	75%
	1	Max

total sulfur dioxide

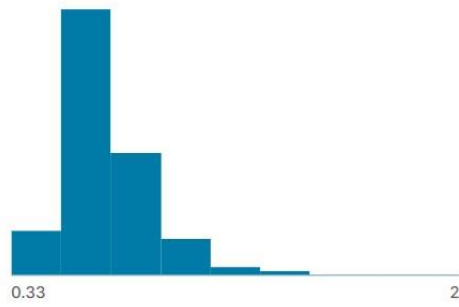
Total sulfur dioxide value



Valid	1143	100%
Mismatched	0	0%
Missing	0	0%
Mean	45.9	
Std. Deviation	32.8	
Quantiles	6	Min
	21	25%
	37	50%
	61	75%
	289	Max

sulphates

Sulphates value



Valid	1143	100%
Mismatched	0	0%
Missing	0	0%
Mean	0.66	
Std. Deviation	0.17	
Quantiles	0.33	Min
	0.55	25%
	0.62	50%
	0.73	75%
	2	Max

KAYNAKÇA

- 1) <https://medium.com/@batubilgili1907.bb/elastic-net-regresyon-modeli-8ce9fcac4ead>
- 2) <https://dergipark.org.tr/tr/download/article-file/1693129>
- 3) <https://esat-akkose1.medium.com/regresyon-ridge-lasso-ve-elastic-net-736b0805fc6>
- 4) https://en.wikipedia.org/wiki/Elastic_net_regularization