

04.06. Linear Discriminant Analysis (LDA – Lineer Diskriminant Analizi)

1936 yılında R. A. Fischer tarafından geliştirilen bir sınıflama metodudur. Basit olmasına rağmen kompleks problemlerde iyi sonuçlar üreten bir modeldir. Bu yazımda bu modeli basit bir örnek üzerinden anlatmaya çalışacağım.

LDA, iyi sınıf (hedefler) arasında en iyi şekilde ayıran değişkenleri lineer bir kombinasyonunu aramaya dayanır. Fisher aşağıdaki score fonksiyonunu tanımlar.

$$Z = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_d x_d$$
$$S(\beta) = \frac{\beta^T \mu_1 - \beta^T \mu_2}{\beta^T C \beta} \quad \text{Score Fonksiyonu}$$

↓

$$S(\beta) = \frac{\bar{Z}_1 - \bar{Z}_2}{Z\text{'nin gruptaki varyansı}}$$

Score fonksiyonuna göre, problem score'u maksimize eden lineer katsayıları tahmin etmedir. Formülasyonu:

$$\beta = C^{-1}(\mu_1 - \mu_2) \quad \text{Model Katsayıları}$$
$$C = \frac{1}{n_1 + n_2} (n_1 C_1 + n_2 C_2) \quad \text{Toplanmış kovaryans matrisi}$$

Where:

β : Lineer model katsayıları

C_1, C_2 : Kovaryans matrisi

μ_1, μ_2 : Ortalama vektörler

En iyi diskriminantı belirleme yolu iki grup arasında Mahalanobis mesafesini hesaplamaktır. Mahalanobis mesafesinin üçten küçük olması yanlış sınıflandırma olasılığını oldukça küçük olduğunu anlamına gelir.

$$\Delta^2 = \beta^T (\mu_1 - \mu_2)$$

Δ : İki grup arasında Mahalanobis farkı

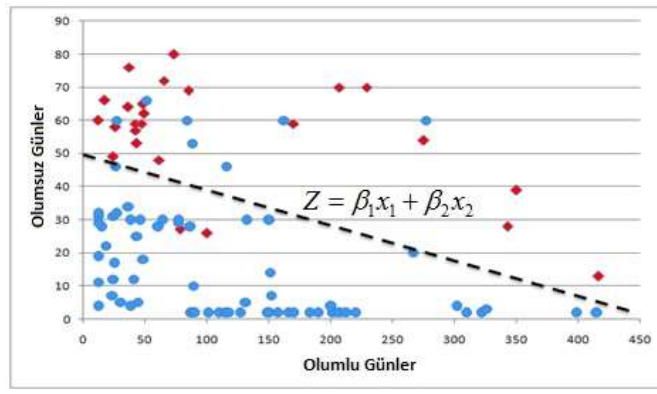
Son olarak, eğer aşağıdaki şart sağlanırsa yeni gelen bir özellik sınıflandırılır.

$$\beta^T \left(x - \left(\frac{\mu_1 + \mu_2}{2} \right) \right) > \log \frac{p(c_1)}{p(c_2)}$$

Konunun daha iyi anlaşılması için örnek bir veri seti üzerinden işlemlerle konuyu anlamaya çalışalım. Ayrıca aşağıdaki excel dosyası üzerinden yapılan formülasyonları ufak bir veri seti üzerinde yapılan örneği inceleyebilirsiniz.

Örnek

Örneğimizde ufak işletmelerin suçlu gün sayısı (OLUMSUZ GÜNLER) ve iş yaptığı gün sayısı (OLUMLU GÜNLER) bir bankanın kredi verip vermeyeceği hakkında veri setini alalım. Bu noktada elimizdeki verilerden en iyi lineer modeli oluşturmaya çalışalım.



Yukarıdaki grafikte kırmızı noktalar bankanın kredi vermediği mavi noktalar ise kredi verdiği bölümleri göstermektedir. Banka genelde 50 gün olumsuz günü olan müşterilerine kredi vermemektedir. (Ama verdikleri de vardır.) Kredi vermesinin sebebi teminat, torpil vs. gibi başka değişkenler olabilir. Biz müşterinin eski bilgilerinden sadece olumlu ve olumsuz gün sayılarını biliyoruz. İlk olarak sınıf olasılıklarını, ortalama vektörleri ve kovaryans matrisini hesaplayalım.

Sınıf (HEDEF)	Gün Sayıları	Olasılık	İstatistik	Olumsuz	Olumlu
Hayır (H)	$n_1=75$	$p(c_1)=0.75$	Ortalama Vektörü (μ)	116.23	16.89
			Covaryans Matrisi (C)	$\begin{bmatrix} 9323 & -607 \\ -607 & 333 \end{bmatrix}$	
Evet (E)	$n_2=25$	$p(c_2)=0.25$	Ortalama Vektörü (μ)	115.04	55.32
			Covaryans Matrisi (C)	$\begin{bmatrix} 14009 & -1053 \\ -1053 & 287 \end{bmatrix}$	

Bu veriler üzerinden, toplanmış kovaryans matrisi ve lineer model katsayılarını hesaplayalım.

$$C = \frac{1}{n_1 + n_2} (n_1 C_1 + n_2 C_2) = \begin{bmatrix} 10495 & -718 \\ -718 & 322 \end{bmatrix}$$

$$\beta = C^{-1}(\mu_1 - \mu_2) = \begin{bmatrix} -0.0095 & -0.1408 \end{bmatrix}$$

$$Z = -0.0095 \text{ BUSAGE} - 0.1408 \text{ DAYSDELQ}$$

Ve son olarak Mahalanobis mesafesini hesaplayalım.

$$\Delta^2 = \beta^T (\mu_1 - \mu_2) = 5.40$$

$$\Delta = 2.32$$

2.32 Mahalanobis mesafesi değerinin 3'ten küçük olması sınıflanan iki grup arasında iyi bir ayrım olduğunu anlamına gelir. Başka bir deyişle sınıflandırma modelimiz iyidir.

Excel dosyasındaki hesaplar için aşağıdaki dosyayı indirin.

[Banka_lda.xlsx \(21,12 kb\)](#)

Referanslar

- <http://www.saedsayad.com/lda.htm>

Makine Öğrenmesi





Search...

- [01. Makine Öğrenmesi](#)
- [02. Supervised ve Unsupervised Learning](#)
- [03. Model Başarısı Değerlendirme – Sınıflandırma](#)
- [04. Sınıflandırma](#)
 - [04.01. K-NN Algoritması](#)
 - [04.02. ZeroR](#)
 - [04.03. OneR](#)
 - [04.04. Naive Bayes Classifier](#)
 - [04.05. Decision Tree \(Karar Ağacı\): ID3 Algoritması – Classification \(Sınıflama\)](#)
 - **[04.06. Linear Discriminant Analysis \(LDA – Lineer Diskriminant Analizi\)](#)**
 - [04.07. Bayesian Ağ](#)
 - [04.07.01. Çok Katmanlı Perceptrons](#)
 - [04.07.02. Destek Karar Makineleri](#)
 - [04.08. Artificial Neural Network \(Yapay Sinir Ağı\)](#)
 - [04.09. Bir Makine Öğrenmesi Uygulaması: Web sayfasından asıl içeriklerin çıkarımı](#)
 - [04.10. Python ile Metin Sınıflandırması Örneği](#)
- [05. Regresyon](#)
- [07. Kümeleme](#)
- [08. Association Rules \(Birliktelik Kuralları\)](#)