

- Naive Bayes Sınıflandırma Algoritmasını anlatın?

1

- Naive Bayes sınıflandırma algoritması, matematikçi Thomas Bayes'in önermiş olduğu olasılık ilkelere temel alan ve veri setindeki özellikler arasındaki bir ilişki olmadığını varsaydığından dolayı
- Naive - (naif, hassas) ismini almış bir supervised (gözetimli, denetimli, eğitilmiş) sınıflandırma algoritmasıdır.

Tahmin işlemi bağımsız değişkenlerin, bağımlı değişkenler üzerindeki etkilerini bir araya getirerek yeni bir durumu sınıflandırmak için yapılır.

Veri setindeki bütün özellikleri eşit sayması Naive Bayes'in yetersizliğidir. Örneğin şeker hastalığı tespit probleminde kişinin kan şekeri ile kişinin boyunun aynı değerde bir özellik olarak görülmesi mantıklı değildir.

Gelişmesi:

$$X = [x_1, x_2, x_3, \dots, x_n] \quad (\text{test verisi})$$

$$C = [c_1, c_2, c_3, \dots, c_k] \quad (\text{sınıf değeri})$$

$$P\left(\frac{c_i}{X}\right) = \frac{P\left(\frac{X}{c_i}\right) \times P(c_i)}{P(X)}$$

Olasılık değerleri
bu formülle hesaplanır

$$\text{arg max}_{c_i} \{ P(X|c_i) P(c_i) \}$$

En büyük olasılık değeri
olan sınıfa atanır.

Eğitim	Yaş	Cinsiyet	Kabul
Orta	Yaşlı	E	Evet
İlk	Genc	E	Hayır
Yüksek	Orta	K	Hayır
Orta	Orta	E	Evet
İlk	Orta	E	Evet
Yüksek	Yaşlı	K	Evet
İlk	Genc	K	Hayır
Orta	Orta	K	Evet

5 EVET
3 HAYIR

TEST (Eğitim = Yüksek, Yaş = Orta, Cinsiyet = Kadın) = ?

Nitelikler	Nitelik Değer	Evet Sayısı	Evet Olanlık	Hayır Sayısı	Hayır Olanlık
Eğitim	İlk	1	1/5	2	2/3
	Orta	3	3/5	0	0
	Yüksek	1	1/5	1	1/3
Yaş	Genc	0	0	2	2/3
	Orta	3	3/5	1	1/3
	Yaşlı	2	2/5	0	0
Cinsiyet	E	3	3/5	1	1/3
	K	2	2/5	2	2/3

C1 = Evet C2 = Hayır

$$P(X|C1) P(C1) \Rightarrow P\left(\frac{X1}{C1}\right) = P\left(\frac{\text{Eğitim} = \text{Yüksek}}{\text{Evett}}\right) = \frac{1}{5}$$

$$= \frac{1}{5} \times \frac{3}{5} \times \frac{2}{5} = \frac{6}{125} = 0.048$$

$$P\left(\frac{X2}{C1}\right) = P\left(\frac{\text{Yaş} = \text{Orta}}{\text{Evett}}\right) = \frac{3}{5}$$

$$P(C1) = P(\text{Evett}) = \frac{5}{8} = 0.625$$

$$P\left(\frac{X3}{C1}\right) = P\left(\frac{\text{Cinsiyet} = \text{K}}{\text{Evett}}\right) = \frac{2}{5}$$

$$= 0.048 \times 0.625 = 0.03$$

$$P(X|C_2) P(C_2) \Rightarrow P\left(\frac{X_1}{C_2}\right) = P\left(\frac{\text{Eğitim=Yüksek}}{\text{Hayır}}\right) = \frac{1}{3} \quad (3)$$

$$= \frac{1}{3} \cdot \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{27} = \boxed{0.074}$$

$$P\left(\frac{X_2}{C_2}\right) = P\left(\frac{\text{Yaş=Orta}}{\text{Hayır}}\right) = \frac{1}{3}$$

$$P(C_2) = P(\text{Hayır}) = \frac{2}{8} = \boxed{0.375}$$

$$P\left(\frac{X_3}{C_2}\right) = P\left(\frac{\text{Cinsiyet=K}}{\text{Hayır}}\right) = \frac{2}{3}$$

$$= 0.074 \times 0.375 = \boxed{0.02775}$$

$$\text{arg max}_{C_i} (P(X|C_i) P(C_i)) = \max\{(0.03), (0.02775)\} = 0.03$$

\Rightarrow EUET ✓

Naive Bayes Sınıflandırma Algoritmasını !!

Avantajları

- Kolay uygulanabilir.
- Gözde durumda başarılı gelsin.
- Metin madenciliğinde başarılıdır.
- Kategorik verilerde sayısal verilere göre daha iyi gelsin.

Dezavantajları

- Her bir ötelemeye eşit değer vermek
- Sistem dinamik olduğunda her defasında eğitim en baştan tekrar yapılır.
- Kategorik verinin olasılık değeri 0 ile Laplace Tahminci'si ile 1 eklemek 0 değeri ortadan kaldırılır.

Kullanım Alanları

Gecede zamanlı tahmin, Çok sınıflı tahmin, Metin sınıflandırma, Spam filtreleme, Duygu analizi, Öneri sistemleri...

