



Bayes Sınıflandırıcılar

- Naïve Bayes Sınıflandırıcı adını İngiliz matematikçi Thomas Bayes'ten (yak. 1701 - 7 Nisan 1761) alır.
- Naïve Bayes Sınıflandırıcı Örüntü tanıma problemine ilk bakışta oldukça kısıtlayıcı görülen bir önerme ile kullanılabilen olasılıkçı bir yaklaşımdır.
- Bu önerme örüntü tanıma da kullanılacak her bir tanımlayıcı nitelik ya da parametrenin istatistik açıdan bağımsız olması gerekliliğidir.



Bayes Sınıflandırıcılar

- Bayes Sınıflayıcı Bayes teoremine göre istatistiksel kestirim yapar.
- Bir örneğin sınıf üyelik olasılığını kestirir.
- *Naïve Bayesian sınıflandırıcı* (simple Bayesian classifier) oldukça başarılı bir sınıflayıcıdır.



Bayes Kuralı

- $p(\mathbf{x}|C_j)$: Sınıf j'den bir örneğin x olma olasılığı
- $P(C_j)$: Sınıf j'nin ilk olasılığı
- $p(\mathbf{x})$: Herhangi bir örneğin x olma olasılığı
- $P(C_j|\mathbf{x})$: x olan bir örneğin sınıf j'den olma olasılığı (son olasılık)

$$P(C_j | \mathbf{x}) = \frac{p(\mathbf{x} | C_j)P(C_j)}{p(\mathbf{x})} = \frac{p(\mathbf{x} | C_j)P(C_j)}{\sum_k p(\mathbf{x} | C_k)P(C_k)}$$



Naïve Bayes sınıflandırıcı

- T öğrenme kümesinde bulunan her örnek n boyutlu uzayda tanımlı olsun, $\mathbf{X} = (x_1, x_2, \dots, x_n)$
- Veri kümesinde m adet sınıf bulunuyor olsun, C_1, C_2, \dots, C_m
- Sınıflamada son olasılığı en büyütme aranır (the maximal $P(C_i|\mathbf{X})$)
- Bayes teoreminden türetilebilir
$$P(C_i|\mathbf{X}) = \frac{P(\mathbf{X}|C_i)P(C_i)}{P(\mathbf{X})}$$
- $P(\mathbf{X})$ olasılığı bütün sınıflar için sabit olduğuna göre, sadece $P(C_i|\mathbf{X}) = P(\mathbf{X}|C_i)P(C_i)$ olasılığı için en büyük değer aranır.

Naïve Bayes sınıflandırıcı



$$P(C_i | \mathbf{X}) = P(\mathbf{X} | C_i) P(C_i)$$

- Eğer bu basitleştirilmiş ifadede bütün özellikler bağımsız ise $P(\mathbf{X} | C_i)$ aşağıdaki şekilde yazılabilir.

$$P(\mathbf{X} | C_i) = \prod_{k=1}^n P(x_k | C_i) = P(x_1 | C_i) \times P(x_2 | C_i) \times \dots \times P(x_n | C_i)$$

- Böylece hesap karmaşıklığı büyük ölçüde azaltılmış olur.

Bayes Sınıflandırıcı için Örnek



Table 10.4 • Data for Bayes Classifier

Magazine Promotion	Watch Promotion	Life Insurance Promotion	Credit Card Insurance	Sex
Yes	No	No	No	Male
Yes	Yes	Yes	Yes	Female
No	No	No	No	Male
Yes	Yes	Yes	Yes	Male
Yes	No	Yes	No	Female
No	No	No	No	Female
Yes	Yes	Yes	Yes	Male
No	No	No	No	Male
Yes	No	No	No	Male
Yes	Yes	Yes	No	Female

Bayes Sınıflandırıcı için Örnek



- Sınıflandırılacak örnek:
 - Magazine Promotion = *Yes*
 - Watch Promotion = *Yes*
 - Life Insurance Promotion = *No*
 - Credit Card Insurance = *No*
 - Sex = ?

Bayes Sınıflandırıcı için Örnek



Table 10.5 • Counts and Probabilities for Attribute Sex

Sex	Magazine Promotion		Watch Promotion		Life Insurance Promotion		Credit Card Insurance	
	Male	Female	Male	Female	Male	Female	Male	Female
Yes	4	3	2	2	2	3	2	1
No	2	1	4	2	4	1	4	3
Ratio: yes/total	4/6	3/4	2/6	2/4	2/6	3/4	2/6	1/4
Ratio: no/total	2/6	1/4	4/6	2/4	4/6	1/4	4/6	3/4

Bayes Sınıflandırıcı için Örnek



- Sex = Male için olasılık hesabı

$$P(\text{sex} = \text{male} \mid E) = \frac{P(E \mid \text{sex} = \text{male}) P(\text{sex} = \text{male})}{P(E)}$$

Bayes Sınıflandırıcı için Örnek



- Sex = Male için koşullu olasılıklar;
 - $P(\text{magazine promotion} = \text{yes} \mid \text{sex} = \text{male}) = 4/6$
 - $P(\text{watch promotion} = \text{yes} \mid \text{sex} = \text{male}) = 2/6$
 - $P(\text{life insurance promotion} = \text{no} \mid \text{sex} = \text{male}) = 4/6$
 - $P(\text{credit card insurance} = \text{no} \mid \text{sex} = \text{male}) = 4/6$
- $P(E \mid \text{sex} = \text{male}) = (4/6) (2/6) (4/6) (4/6) = 8/81$

$$P(\text{sex} = \text{male} \mid E) \approx (8/81) (6/10) / P(E)$$

$$P(\text{sex} = \text{male} \mid E) \approx 0,0593 / P(E)$$

Bayes Sınıflandırıcı için Örnek



- Sex = Female için olasılık hesabı

$$P(\text{sex} = \text{female} \mid E) = \frac{P(E \mid \text{sex} = \text{female}) P(\text{sex} = \text{female})}{P(E)}$$

Bayes Sınıflandırıcı için Örnek



- Sex = Female için koşullu olasılıklar;
 - $P(\text{magazine promotion} = \text{yes} \mid \text{sex} = \text{female}) = 3/4$
 - $P(\text{watch promotion} = \text{yes} \mid \text{sex} = \text{female}) = 2/4$
 - $P(\text{life insurance promotion} = \text{no} \mid \text{sex} = \text{female}) = 1/4$
 - $P(\text{credit card insurance} = \text{no} \mid \text{sex} = \text{female}) = 3/4$
- $P(E \mid \text{sex} = \text{female}) = (3/4) (2/4) (1/4) (3/4) = 9/128$
$$P(\text{sex} = \text{female} \mid E) \approx (9/128) (4/10) / P(E)$$
$$P(\text{sex} = \text{female} \mid E) \approx 0,0281 / P(E)$$

Bayes Sınıflandırıcı için Örnek



- $P(\text{sex} = \text{male} \mid E) \approx 0,0593 / P(E)$
- $P(\text{sex} = \text{female} \mid E) \approx 0,0281 / P(E)$

Bayes sınıflayıcı $0,0593 > 0,0281$ olduğu için E davranışını gösteren kart sahibi erkektir.

Bayes Sınıflandırıcı için Örnek



1-Gerçekleştireceğimiz analizler için kullanacağımız kütüphanemizi projemize dahil edelim

```
# ÇALIŞMA ÜZERİNDE KULLANILACAK KUTUPHANELERİ DAHİL EDELİM.  
import pandas as pd
```

2-Çalışmamıza ilgili verilerimizi dahil edelim ve önizleme gerçekleştirelim;

```
# İLGİLİ VERİ SETİNİ ÇALIŞMAMIZA DAHİL EDELİM.  
Iris_Data = pd.read_csv("D:\Yedekleme\Iris.csv")
```

Iris_Data

	Id	SepalLengthCm	SepalWidthCm	PetalLengthCm	PetalWidthCm	Species
0	1	5.1	3.5	1.4	0.2	Iris-setosa
1	2	4.9	3.0	1.4	0.2	Iris-setosa
2	3	4.7	3.2	1.3	0.2	Iris-setosa
3	4	4.6	3.1	1.5	0.2	Iris-setosa
4	5	5.0	3.6	1.4	0.2	Iris-setosa
...
145	146	6.7	3.0	5.2	2.3	Iris-virginica
146	147	6.3	2.5	5.0	1.9	Iris-virginica
147	148	6.5	3.0	5.2	2.0	Iris-virginica
148	149	6.2	3.4	5.4	2.3	Iris-virginica
149	150	5.9	3.0	5.1	1.8	Iris-virginica

150 rows × 6 columns

Bayes Sınıflandırıcı için Örnek



3-Bağımlı değişkenimizi analiz gerçekleştirmek için bir parametreye atayalım; (*Species: Tür-Cins*)

```
# BAGIMLI DEGISENIMIZI BELIRLEMIS OLDUGUMUZ BIR DEGISKENE ATADIK.  
# SPECIES: TUR-CESIT-CINS KAVRAMI.  
Data_Species = Iris_Data.iloc[:, -1:].values
```

4-Veri kümemizi test ve eğitim/öğrenme kümeleri olarak ikiye ayıralım;

```
# VERI KUMEMIZI TEST VE EGITIM VERISI OLARAK BOLUYORUZ.  
from sklearn.model_selection import train_test_split  
x_train, x_test, y_train, y_test = train_test_split(Iris_Data.iloc[:, 1:-1], Data_Species, test_size=0.35, random_state=0)
```

5-Çalışmamıza analiz gerçekleştirmek için GaussianNB sınıfını dahil edelim; (Sklearn)

```
# CALISMAMIZA GaussianNB SINIFINI DAHIL ETTIK.  
# CALISMA YAPINIZI ETKILEYECEK 5 FARKLI NAIVE BAYES SINIFI MEVCUTTUR.  
# HEPSI ILGILI KUTUPHANE ICERISINDE YER ALMAKTADIR.  
# BUNLAR;  
# GaussianNB, BernoulliNB, MultinomialNB, CategoricalNB, ComplementNB.  
# GERCEKLESTIRECEGINIZ KURGUYA GORE BU 5 SINIFTAN BIRINI TERCIH EDEBILIRSINIZ.  
# BU SINIFLAR DOGRUDAN MODELIN BASARISINI ETKILEMEKTEDIR.  
# EK DETAYLAR ICIN; https://scikit-learn.org/stable/modules/classes.html#module-sklearn.naive\_bayes  
from sklearn.naive_bayes import GaussianNB
```

Bayes Sınıflandırıcı için Örnek



6-Analiz için GaussianNB sınıfından bir nesne türetelim/üretelim;

```
# ANALIZ ICIN GaussianNB SINIFINDAN BIR NESNE URETTIK.  
GaussianNB_Class = GaussianNB()
```

7-Oluşturmuş olduğumuz nesnemiz üzerinde verimizi eğitelim; (*Training Process*)

```
# YAPIMIZI EGITMEYE BASLIYORUZ.  
GaussianNB_Class.fit(x_train, y_train.ravel())
```

8-Eğitmiş olduğumuz yapımız üzerinde tahminleme gerçekleştirelim; (*Predict Process – Prediction*)

```
# TEST VERI KUMEMIZ UZERINDE TAHMINLEME GERCEKLESTIRIYORUZ.  
Result = GaussianNB_Class.predict(x_test)
```

9-Tahminleme yapımız üzerinde başarı oranı hesabı adına karmaşıklık matrisi inşaa edelim ve görüntüleyelim;
(*Confusion Matrix – Karmaşıklık/Karışıklık Matrisi*)

```
# SISTEM ANALIZI ICIN KARMASIKLIK MATRISI INSAA EDIYORUZ.  
from sklearn.metrics import confusion_matrix  
Confusion_Matrix = confusion_matrix(y_test, Result)  
print(Confusion_Matrix)
```

```
[[16  0  0]  
 [ 0 21  0]  
 [ 0  2 14]]
```


Bayes Sınıflandırıcı için Örnek



10-Matrisimiz üzerinden başarı oranımızı hesaplayalım;

```
# SISTEM BASARI ORANI HESAPLAMASI;  
from sklearn.metrics import accuracy_score  
Accuracy = accuracy_score(y_test, Result)  
print(Accuracy)
```

0.9622641509433962

```
[ [16  0  0]  
  [ 0 21  0]  
  [ 0  2 14]
```

$16+21+14+2=53$

$16+21+14=51$

$51/53=0,9622...$

Algoritmamız ile %96 oranında bir başarı sonucu çıktısı aldık.

Bu çıktı sonuçlarını GaussianNB sınıfı dışındaki diğer analiz sınıfları ile karşılaştırabilirsiniz.

(GaussianNB yerine; BernoulliNB, MultinomialNB, CategoricalNB, ComplementNB sınıflarını baz alabilirsiniz.)

Naive Bayes algoritmasına yönelik basit veri seti üzerinden kısa bir analiz gerçekleştirdik.

Bayes Sınıflandırıcı için Örnek



Günlük Hayatta Algoritma Uygulama Alanları

Günlük hayatta Naive Bayes algoritması esnek, basit uyarlanabilir ve hızlı analiz yapısı sayesinde genel olarak tüm;

- Gerçek ve eş zamanlı tüm analiz projelerinde, (*Real Time Computing Project*)
- Metin Sınıflandırması problem ve projelerinde, (*Document-Text Classification*)
- Tavsiye Sistemleri problem ve projelerinde, (*Recommendation System*)
- Duygu Analizi yörengeli sınıflandırma problem ve projelerinde (*Sentiment Analysis*)

yer alıp, kullanılabilmektedir.

Avantaj Dezavantaj

- Basit ve kolay uygulanabilir bir yapısı bulunmaktadır,
- Yapısal olarak problemlı (dengesiz) veri kümelerinde de kullanılabılır,
- Hızlı yapıda analiz gerçekteştirmesinden dolayı gerçekte zamanlı/eş zamanlı projeler üzerinde kullanılabılır,
- Özelleştirilebilir yüksek yapıdaki veriler üzerinde iyi çalışmaktadır.
- Veri kümesi içerisinde yer alan değışkenler arasında modelleme gerçekteştirilmesine uygun değildir,
- Sıkça *Zero Probability* problemi ile karşı karşıya kalınmaktadır,
- Veri kümesi üzerindeki tüm özelliklerin bağımsız olduğunu varsaymaktadır; teorik olarak harika olduğu düşünölse de, gerçekte hayatta bu yapıdaki dizi üzerinde bağımsız özellik bulamazsınız.