



Naive Bayes

Naive Bayes :

- Naive Bayes Algoritması, sınıflandırma görevlerinde kullanılan en önemli makine öğrenmesi algoritmalarından biridir. Bu algoritma, değişkenler arasındaki bağımsızlık varsayımı ile Bayes Teoremi'ne dayanır. Başka bir deyişle, Naive Bayes algoritması, bir özelliğin bir sınıfta bulunmasının diğer herhangi bir özelliğin bulunmasıyla ilişkili olmadığını varsayar. Algoritma, belirli bir sınıfla ilişkili özelliklerin olasılıklarına dayanarak, yeni bir veri noktasının belirli bir sınıfa ait olma olasılığını hesaplayarak tahminlerde bulunur. Algoritma, özelliklerin birbirinden bağımsız olduğunu varsaydığı için "Naive (Saf)" olarak adlandırılır.

- Örneğin, bir meyveyi kırmızı, yuvarlak ve çapı yaklaşık 3 inç olan bir elma olarak sınıflandırabilir. Bu özellikler birbirine bağımlı olsa veya diğer özelliklerin varlığına bağlı olsa bile, Naive Bayes sınıflandırıcısı bu özelliklerin her birini bağımsız olarak ele alarak bu meyvenin elma olma olasılığını tahmin eder.
- **Naive Bayes'in Başlıca Avantajları:**
- **Basitlik:** Kullanımı ve anlaşılması kolaydır.
- **Hız:** Hızlı çalışır.
- **Doğruluk:** Özellikle metin sınıflandırma görevleri ve çok büyük veri kümeleri için oldukça doğrudur.
- **Aşırı Uyuma Direnç:** Diğer makine öğrenmesi algoritmalarına göre aşırı uyuma daha az eğilimlidir.
- **Dezavantajları:**
- **Bağımlı Veriler:** Veri özellikleri birbirine çok bağımlı olduğunda veya çok fazla özellik olduğunda, diğer daha karmaşık algoritmalar kadar iyi performans göstermeyebilir.
- **Bayes Teoremi:** Naive Bayes algoritması, $P(c | x)$ olasılığını $P(c)$, $P(x)$ ve $P(x | c)$ kullanarak hesaplamak için Bayes Teoremi'ni kullanır. Aşağıdaki denkleme bakın:

$$P(c | x) = \frac{P(x | c)P(c)}{P(x)}$$

Diagram labels: Likelihood (points to $P(x | c)$), Class Prior Probability (points to $P(c)$), Posterior Probability (points to $P(c | x)$), Predictor Prior Probability (points to $P(x)$)

$$P(c | X) = P(x_1 | c) \times P(x_2 | c) \times \dots \times P(x_n | c) \times P(c)$$

$P(c | x)$: Verilen bir özellik kümesi (x) için sonraki sınıf olasılığı.
 $P(c)$: Sınıfın öncül olasılığı. $P(x | c)$: Belirli bir sınıf için özelliklerin olasılığı. $P(x)$: Özelliklerin öncül olasılığı.

Algoritmanın çalışma yöntemi :

- Bu algoritmanın çalışma yöntemini anlamak için aşağıdaki örneğe bakalım. Bu örnekte, hava durumuna göre futbol oynama olasılığını tahmin etme işlemi göreceğiz.
- Aşama 1: Veri setini tablo haline getirme. Aşama 2: Olasılık tablosunu oluşturma. Örneğin, bulutlu hava olasılığı = 0.29 ve oynama olasılığı = 0.64 gibi olasılıkları bulma. Aşama 3: Naive Bayes denklemi kullanarak her kategori için sonraki olasılığı hesaplama. En yüksek sonraki olasılığa sahip kategori tahmin edilen sonucu oluşturacaktır.

| Weather | Play |
|----------|------|
| Sunny | No |
| Overcast | Yes |
| Rainy | Yes |
| Sunny | Yes |
| Sunny | Yes |
| Overcast | Yes |
| Rainy | No |
| Rainy | No |
| Sunny | Yes |
| Rainy | Yes |
| Sunny | No |
| Overcast | Yes |
| Overcast | Yes |
| Rainy | No |

| Frequency Table | | |
|-----------------|----|-----|
| Weather | No | Yes |
| Overcast | | 4 |
| Rainy | 3 | 2 |
| Sunny | 2 | 3 |
| Grand Total | 5 | 9 |

| Likelihood table | | |
|------------------|-------|-------|
| Weather | No | Yes |
| Overcast | | 4 |
| Rainy | 3 | 2 |
| Sunny | 2 | 3 |
| All | 5 | 9 |
| | =5/14 | =9/14 |
| | 0.36 | 0.64 |

- Problemi çözmek için yukarıda tartışılan yöntemi kullanabiliriz, bu nedenle:
- $P(\text{Evet} \mid \text{Güneşli}) = P(\text{Güneşli} \mid \text{Evet}) * P(\text{Evet}) / P(\text{Güneşli})$
- Burada,
- $P(\text{Güneşli} \mid \text{Evet}) = 3/9 = 0.33$
- $P(\text{Güneşli}) = 5/14 = 0.36$
- $P(\text{Evet}) = 9/14 = 0.64$
- Dolayısıyla,
- $P(\text{Evet} \mid \text{Güneşli}) = 0.33 * 0.64 / 0.36 = 0.60$
- 0.6 değeri oldukça yüksek bir olasılık değeridir.
- Naive Bayes algoritması, farklı özelliklere dayanarak farklı sınıfların olasılıklarını tahmin etmek için benzer bir yöntemi kullanır. Bu algoritma genellikle metin sınıflandırması ve çoklu sınıflı problemlerde kullanılır.

- **Naive Bayes'in farklı türleri vardır**, ancak en yaygın olanları Gauss Naive Bayes, Çoklu Naive Bayes ve Bernoulli Naive Bayes sınıflayıcılarıdır. Sınıflayıcı seçimi, analiz edilen veri türüne ve veriler hakkında yapılabilecek varsayımlara bağlıdır.
- **Gaussian Naive Bayes (GNB)**: Gauss Naive Bayes sınıflayıcı, özelliklerin sürekli ve normal dağıldığı varsayımını yapar. Genellikle, özelliklerin Gauss veya normal dağılıma sahip olduğu sınıflandırma problemlerinde kullanılır, örneğin doğal dil işlemede, özellikler kelime sayısını veya frekanslarını temsil edebilir.
- **Multinomial Naive Bayes (MNB)**: Çoklu Naive Bayes sınıflayıcısı, ayrık sayı verileri için kullanılır, örneğin metin sınıflandırmasında kelime sayısı gibi. Özelliklerin bağımsız olduğu ve çoklu dağılımı izlediği varsayılır. Doküman sınıflandırması, spam filtreleme ve duygu analizi gibi alanlarda sıkça kullanılır.
- **Complement Naive Bayes (CNB)**: Complement Naive Bayes sınıflayıcısı, dengesiz veri kümeleri ile başa çıkmak için geliştirilmiş bir Multinomial Naive Bayes sınıflayıcısıdır. Geleneksel Multinomial Naive Bayes'de, algoritma her sınıftaki her özelliğin tekrarını tahmin eder ve dengesiz veri setleriyle olası önyargılı sonuçlara yol açabilir. CNB'de, algoritma her sınıftaki her özelliğin tamamlayıcısının tekrarını tahmin eder, bu da baskın sınıfın etkisini azaltır ve dengesiz veri setlerinde daha iyi performans sağlayabilir.
- **Bernoulli Naive Bayes (BNB)**: Bernoulli Naive Bayes sınıflayıcısı, ikili veriler için kullanılır, yani özellikler ya var ya da yoktur. Özelliklerin bağımsız olduğu ve Bernoulli dağılımını takip ettiği varsayılır. Metin sınıflandırmasında kullanılır, belirli kelimelerin veya ifadelerin varlığı veya yokluğu özellikler olarak kullanıldığında.
- **Categorical Naive Bayes (CatNB)**: Categorical Naive Bayes sınıflayıcısı, kategorik veriler için tasarlanmış bir çoklu sınıflandırma algoritmasıdır. Her bir özelliğin sınırlı bir kategori kümesinden birini alabileceği varsayılır ve her kategori tekrarının her sınıfta tahmin edilmesi için olasılıklar atanır. Metin sınıflandırmasında konu sınıflandırması veya duygu analizi gibi alanlarda yaygın olarak kullanılır.