

Naive Bayes

Naive Bayes, denetimli öğrenme problemleri için kullanılan bir makine öğrenmesi algoritmasıdır. Sınıflandırma ve olasılık tahmini gibi çeşitli problemlerde yaygın olarak kullanılır. Naive Bayes algoritması, Bayes teoremi temel alınarak çalışır.

Bayes teoremi, bir olayın meydana gelme olasılığını hesaplamanın bir yoludur. Bayes teoremi, olayların rastgele değişkenler olarak modellendiği ve belirli bir sonucun gerçekleşme olasılığının, bu olayların gözlemlendiği bir veri kümesi temelinde hesaplandığı birçok alanda kullanılır.

Naive Bayes algoritması, Bayes teoremini temel alır ve "naive" (saf) olarak adlandırılır, çünkü sınıf etiketlerinin birbirinden bağımsız olduğunu ve her bir özellik için etiketlerin birlikte gözlemlenebilirliğini varsayar. Bu varsayım, modelin basitleştirilmesini sağlar ve hesaplamanın daha hızlı olmasını sağlar.

Naive Bayes algoritması, bir örneğin bir sınıfa ait olma olasılığını tahmin etmek için Bayes teoreminden yararlanır. Verilen bir girdi örneği için, her bir sınıfın olasılığını hesaplamak için özelliklerin olasılıklarını kullanır. Ardından, bu olasılıkları kullanarak, örneğin her bir sınıfa ait olma olasılığını hesaplar ve en yüksek olasılığa sahip sınıfı tahmin eder.

Özetle, Naive Bayes algoritması, özelliklerin birbirinden bağımsız olduğunu varsayarak, Bayes teoreminden yararlanarak sınıflandırma yapar. Bu basit ve etkili yaklaşım, özellikle metin sınıflandırması gibi alanlarda popülerdir.

Naive Bayes algoritmasının formülü, Bayes teoremini kullanarak bir örneğin bir sınıfa ait olma olasılığını hesaplamak için şu şekildedir:

$$P(y|x_1, x_2, \dots, x_n) = \frac{P(x_1|y) \times P(x_2|y) \times \dots \times P(x_n|y) \times P(y)}{P(x_1) \times P(x_2) \times \dots \times P(x_n)}$$

Burada:

- $P(y|x_1, x_2, \dots, x_n)$: Özellikler x_1, x_2, \dots, x_n verildiğinde sınıf y olasılığı (Posterior)
- $P(x_i|y)$: Sınıf y verildiğinde özellik x_i olasılığı (Likelihood)
- $P(y)$: Sınıfın öncül olasılığı (Prior)
- $P(x_i)$: Özellik x_i olasılığı (Evidence)

Uygulama

load_digits veri seti kullanılmıştır

Doğruluk: 0.8472222222222222