

Naive Bayes

Üolasilığa dayalı bir sınıflandırma algoritmasıdır. Temel olarak, bir örneğin belirli bir sınıfa ait olma olasılığını tahmin etmek için Bayes teoremini kullanır. "Naive" (saf) terimi, bu algoritmanın her özelliğın birbirinden bağımsız olduğunu varsaymasıyla ilgilidir, yani bir özelliğın değeri diğer özelliklerin değeri hakkında hiçbir bilgi sağlamaz.

Temel İlkeler

1. Bayes Teoremi:

- Bir olayın gerçekleşme olasılığı hakkında bilgi sahibi olduğumuzda, bu olayın neden olduğu koşullarda diğer olayların olasılıklarını tahmin etmek için kullanılır.
- Bayes teoremi, $P(A|B) = P(B|A) * P(A) / P(B)$ formülüyle ifade edilir.

2. Bağımsızlık Varsayımı:

- Naive Bayes, her özelliğın sınıf etiketi üzerinde bağımsız olduğunu varsayar. Yani, özellikler arasındaki ilişkiyi yok sayar.

3. Sınıflandırma:

- Bir örneğın sınıfını belirlemek için, her sınıfın koşullu olasılığı hesaplanır ve en yüksek olasılığa sahip sınıf seçilir.

Naive Bayes Algoritması Adımları

1. Veri Hazırlığı:

- Veri seti, etiketli örneklerden oluşmalıdır. Her bir örnek, özellikler ve bir sınıf etiketi içermelidir.

2. Eğitim:

- Sınıf etiketleri üzerindeki koşullu olasılıklar hesaplanır.
- Her bir özelliğın belirli bir sınıf altında belirli bir değere sahip olma olasılığı hesaplanır.

3. Tahmin:

- Yeni bir örnek geldiğinde, her sınıfın olasılığı hesaplanır ve en yüksek olasılığa sahip sınıf tahmin olarak seçilir.

Örnek: E-posta Sınıflandırması

Örneğin, bir e-postanın spam olup olmadığını belirlemek için Naive Bayes kullanabiliriz:

1. Eğitim Verisi:

- Eğitim verisi, spam ve spam olmayan e-postalardan oluşmalıdır.
- Her e-posta, kelime frekansları gibi özelliklerle temsil edilir.

2. Eğitim:

- Spam ve spam olmayan e-postaların özelliklerine dayalı olarak sınıf etiketleri üzerinde koşullu olasılıklar hesaplanır.

3. Tahmin:

- Yeni bir e-posta geldiğinde, her sınıfın olasılığı hesaplanır ve en yüksek olasılığa sahip sınıf tahmin olarak seçilir.

Avantajlar ve Dezavantajlar

Avantajlar:

- Basit ve hızlı: Eğitim ve tahmin süreçleri hızlıdır ve küçük veri setleri için uygundur.
- İyi performans: Doğru kullanıldığında, sınıflandırma performansı iyidir ve çok sayıda özellikle başa çıkabilir.

Dezavantajlar:

- Bağımsızlık varsayımı: Özellikler arasındaki gerçek ilişkileri dikkate almaz, bu nedenle gerçek dünya verileri için bazen yetersiz olabilir.
- Sıfır olasılıklar: Eğitim veri setinde belirli bir sınıfa ait olmayan bir özellik değeri görüldüğünde sıfır olasılıkla sonuçlanabilir, bu da tahminlerde yanlışlıkla yol açabilir.

Naive Bayes, özellikle doğal dil işleme, spam filtreleme, duygu analizi gibi alanlarda sıkça kullanılan bir algoritmadır.