

Numpy Korelasyon

Korelasyon Nedir?

Korelasyon, iki veya daha fazla **değişken** arasındaki **istatistiksel ilişkiyi** ifade eder. Basitçe söylemek gerekirse, bir değişken değiştiğinde diğer değişkenin ne ölçüde ve ne yönde değişme eğiliminde olduğunu gösterir.

- **Yön:**

- **Pozitif Korelasyon:** İki değişken aynı yönde hareket etme eğilimindedir. Biri artarken diğeri de artar (veya biri azalırken diğeri de azalır). Örnek: Boy uzunluğu ile kilo arasında genellikle pozitif korelasyon vardır (daha uzun insanlar genellikle daha ağırdır).
- **Negatif Korelasyon:** İki değişken zıt yönlerde hareket etme eğilimindedir. Biri artarken diğeri azalır. Örnek: Bir ürünün fiyatı ile talep edilen miktar arasında genellikle negatif korelasyon vardır (fiyat arttıkça talep azalır).
- **Sıfır (veya Sıfıra Yakın) Korelasyon:** İki değişken arasında belirgin bir doğrusal ilişki yoktur. Birindeki değişim, diğeriindeki değişimi tahmin etmede pek yardımcı olmaz. Örnek: Ayakkabı numarası ile zeka seviyesi arasında muhtemelen sıfır korelasyon vardır.

- **Güç:**

Korelasyonun gücü genellikle

korelasyon katsayısı (en yaygın Pearson korelasyon katsayısıdır, 'r' ile gösterilir) ile ölçülür. Bu katsayı **1 ile +1** arasında bir değer alır:

- **+1:** Mükemmel pozitif doğrusal ilişki.
- **1:** Mükemmel negatif doğrusal ilişki.
- **0:** Doğrusal ilişki yok.
- 0 ile +1 arasındaki değerler pozitif ilişkinin gücünü (1'e yaklaştıkça güçlenir), 0 ile -1 arasındaki değerler negatif ilişkinin gücünü (-1'e yaklaştıkça güçlenir) gösterir. Örneğin, 0.8 güçlü bir pozitif ilişkiyi, -0.2 zayıf bir negatif ilişkiyi gösterir.

- **Önemli Uyarı: Korelasyon Nedensellik Değildir!**

İki değişken arasında güçlü bir korelasyon olması, birinin diğerine **neden olduğu** anlamına gelmez. İlişki tesadüfi olabilir veya her iki değişkeni

de etkileyen üçüncü bir faktör (gizli değişken) olabilir. Klasik örnek: Yaz aylarında dondurma satışları ile suç oranları arasında pozitif korelasyon vardır. Ancak dondurma yemek suça neden olmaz (veya tersi). Her ikisi de sıcak hava nedeniyle artar.

NumPy'daki `corrcoef` Ne İşe Yarar?

`numpy.corrcoef()`, verilen veriler için **Pearson korelasyon katsayılarını** hesaplayan bir NumPy fonksiyonudur. Hesaplama sonucunda bir **korelasyon matrisi** döndürür.

- **Girdi:** Genellikle 1 boyutlu veya 2 boyutlu dizi benzeri (array-like) yapılar alır.
 - Eğer iki 1 boyutlu dizi verirsiniz (örn: `np.corrcoef(x, y)`), bu iki değişken arasındaki korelasyon matrisini hesaplar.
 - Eğer 2 boyutlu bir dizi verirsiniz (örn: `np.corrcoef(A)`), varsayılan olarak her bir **satırı** ayrı bir değişken olarak kabul eder ve bu değişkenler arasındaki korelasyonları hesaplar.
- **Çıktı (Korelasyon Matrisi):** Döndürdüğü matris şu özelliklere sahiptir:
 - Karedir (satır ve sütun sayısı eşittir).
 - Simetriktir (`C[i, j] == C[j, i]`).
 - Köşegen elemanları (sol üstten sağ alta) her zaman **1**'dir. Çünkü bir değişkenin kendisiyle korelasyonu mükemmeldir (+1).
 - Köşegen dışındaki eleman `C[i, j]`, `i`. değişken ile `j`. değişken arasındaki korelasyon katsayısını gösterir.
- **`rowvar` Parametresi:** Bu parametre önemlidir.
 - `rowvar=True` (Varsayılan): Girdi dizisindeki **satırların** değişkenleri, sütunların ise gözlemleri temsil ettiği varsayılır.
 - `rowvar=False`: Girdi dizisindeki **sütunların** değişkenleri, satırların ise gözlemleri temsil ettiği varsayılır. Veri analizi (özellikle Pandas DataFrame'ler gibi) bağlamında genellikle sütunlar değişkenleri tuttuğu için `rowvar=False` sıkça kullanılır.

Örnek:

Python

```
import numpy as np
```

```
# İki değişken (biraz ilişkili)
```

```
x = np.array([1, 2, 3, 4, 5, 6])
y = np.array([2, 3, 5, 7, 8, 10]) # x arttıkça y de artma eğiliminde
```

```
# Korelasyon katsayıları matrisini hesapla
korelasyon_matrisi = np.corrcoef(x, y)
```

```
print("x:", x)
print("y:", y)
print("\nKorelasyon Matrisi:")
print(korelasyon_matrisi)
```

Çıktı:

```
x: [1 2 3 4 5 6]
y: [ 2  3  5  7  8 10]
```

```
Korelasyon Matrisi:
[[1.      0.97629516] # Matris[0, 0]: x'in kendisiyle korelasyonu (1)
 [0.97629516 1.      ] # Matris[0, 1]: x ile y arasındaki korelasyon (yaklaşık 0.976)
 [                   ] # Matris[1, 0]: y ile x arasındaki korelasyon (aynı)
 [                   ] # Matris[1, 1]: y'nin kendisiyle korelasyonu (1)
```

Bu örnekte, `corrcoef` fonksiyonu x ve y arasında yaklaşık 0.976 değerinde güçlü bir pozitif korelasyon olduğunu hesaplamıştır. Matrisin köşegenleri beklendiği gibi 1'dir.