

# Özellik Vektörleri

Makine öğrenmesi özellik vektörleri (feature vectors), bir makine öğrenmesi modelinin eğitimi ve tahmin yapması için kullandığı veriyi temsil eden çok boyutlu matematiksel yapılar olarak tanımlanabilir. Özellik vektörleri, bir gözlem ya da veri noktasının çeşitli özelliklerini veya niteliklerini sayısal olarak ifade eder. Her bir özellik vektörünün bileşenleri, belirli bir gözlemin özellik değerlerini içerir.

## ÖRNEK1

Örneğin, bir ev fiyatını tahmin etmek için bir model oluşturduğunuzu düşünelim. Bu durumda, her ev için aşağıdaki gibi özellikler belirleyebiliriz:

- Metrekare
- Oda sayısı
- Banyo sayısı
- Yaş
- Konum

Bu özelliklerin her biri birer sayısal değere dönüştürülür ve bir özellik vektörü oluşturulur. Örneğin, bir evin özellik vektörü şöyle görünebilir:

Özellik Vektörü=[120,3,2,5,1]\text{Özellik Vektörü} = [120, 3, 2, 5, 1]

Burada:

- 120: Metrekare (m<sup>2</sup>)
- 3: Oda sayısı
- 2: Banyo sayısı
- 5: Evin yaşı (yıl)
- 1: Konum (belli bir konum kodu ile ifade edilmiştir)

Bu vektör, makine öğrenmesi modelinin ev fiyatını tahmin etmek için kullanacağı giriş verisini temsil eder.

## ÖRNEK2

Bu sefer, bir araba fiyatı tahmin modeli oluşturmak istediğimizi varsayalım. Bu durumda, arabalarla ilgili bazı özellikler belirleyebiliriz:

- Yaş (yıl)
- Kilometre (km)
- Motor Hacmi (cc)
- Yakıt Türü (benzin, dizel vb., sayısal olarak kodlanmış)
- Marka (sayısal olarak kodlanmış)

Bu özellikleri kullanarak her araba için bir özellik vektörü oluşturabiliriz. Örneğin, bir arabanın özellik vektörü şöyle olabilir:

Özellik Vektörü=[5,60000,2000,1,3]\text{Özellik Vektörü} = [5, 60000, 2000, 1, 3]Özellik Vektörü=[5,60000,2000,1,3]

Burada:

- 5: Arabanın yaşı (5 yıl)
- 60000: Kilometresi (60.000 km)
- 2000: Motor hacmi (2000 cc)
- 1: Yakıt türü (1, benzin için)
- 3: Marka (3, belirli bir marka için kodlanmış)

Bu vektör, makine öğrenmesi modeline beslenir ve model, bu özellikleri kullanarak arabanın fiyatını tahmin eder.

Aşağıdaki kodda bu örneği temsilen bir kod bulunmaktadır.

## KOD2

```
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error

# Basit veri kümesi oluşturma
data = {
    'mileage': [50000, 30000, 40000, 80000, 60000],
    'age': [5, 3, 4, 8, 6],
    'price': [15000, 20000, 18000, 12000, 14000]
}
df = pd.DataFrame(data)

# Özellikler (X) ve hedef değişken (y)
X = df[['mileage', 'age']]
y = df['price']

# Veriyi eğitim ve test setlerine ayırma
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

# Model oluşturma ve eğitme
model = LinearRegression()
model.fit(X_train, y_train)
```

```
# Tahmin yapma
y_pred = model.predict(X_test)

# Model performansını değerlendirme
mse = mean_squared_error(y_test, y_pred)
print(f'Mean Squared Error: {mse}')

# Yeni bir veri noktası için tahmin yapma
new_car = {'mileage': 35000, 'age': 4}
new_car_df = pd.DataFrame([new_car])
new_price = model.predict(new_car_df)
print(f'Tahmin edilen fiyat: ${new_price[0]:,.2f}')
```

### Not: Kodlara harici dosyadan da ulaşabilirsiniz

Bu kodda:

1. Basit bir veri kümesi oluşturduk.
2. Özellikler (mileage ve age) ve hedef değişken (price) belirledik.
3. Veriyi eğitim ve test setlerine ayırdık.
4. Basit bir lineer regresyon modeli eğittik.
5. Test verileri üzerinde tahmin yaptık ve modelin hata metriğini hesapladık.
6. Yeni bir araba için fiyat tahmini yaptık.

Bu basit örnek, temel bir veri kümesi ve özelliklerle araba fiyat tahmininin nasıl yapılacağını gösterir.

### ÖRNEK3

Başka bir örnek olarak, bir spam e-posta tespiti modelini düşünebiliriz. Bu model, bir e-postanın spam olup olmadığını tahmin etmeye çalışır. Özellikler şunlar olabilir:

- E-postanın uzunluğu (kelime sayısı)
- E-postada belirli anahtar kelimelerin bulunma sıklığı (örneğin, "ücretsiz", "kazan", "hemen" gibi)
- E-posta başlığındaki ünlem işaretlerinin sayısı
- Gönderici e-posta adresinin güvenilirlik skoru
- E-postanın gönderildiği saat dilimi (gün içi, gece vb.)

Özellik vektörü şöyle olabilir:

Özellik Vektörü=[250,5,2,0.7,1] \text{Özellik Vektörü} = [250, 5, 2, 0.7, 1]

Burada:

- 250: E-postanın uzunluğu (250 kelime)
- 5: Belirli anahtar kelimelerin bulunma sıklığı (5 kez)

- 2: Başlıkta iki ünlem işareti bulunması
- 0.7: Gönderici adresinin güvenilirlik skoru (0 ile 1 arasında)
- 1: E-posta gece saatlerinde gönderilmiş (örneğin, 1 geceyi temsil ediyor)

Bu vektörler, makine öğrenmesi modeline e-posta verisi olarak beslenir ve model, bu verileri kullanarak e-postanın spam olup olmadığını tahmin eder.

Bu örnekler, makine öğrenmesi modellerinde kullanılan özellik vektörlerinin ne olduğunu ve nasıl oluşturulduğunu göstermektedir.

Özellik vektörlerinin seçimi ve doğru bir şekilde oluşturulması, makine öğrenmesi modellerinin başarısı için kritik öneme sahiptir. Özellik mühendisliği (feature engineering), bu süreçte önemli bir rol oynar ve genellikle veri temizleme, normalizasyon, ölçekleme ve dönüştürme gibi adımları içerir.

Özetle, özellik vektörleri, makine öğrenmesi modellerine beslenen ve her gözlemi sayısal bir şekilde ifade eden veri kümeleridir.