

## ÖDEV 1

SELEN ERDOĞAN

Selenerdogan2019@gtu.edu.tr

Elektronik Mühendisliği Bölümü, GTÜ, Kocaeli, Türkiye

### I. GİRİŞ

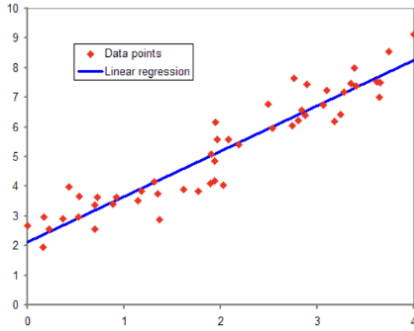
Bu rapor, "samples.txt" dosyasındaki verilerin bir modelle modellenmesi ve modelin parametrelerinin tahmin edilmesi işlemini ele almaktadır. Ödevin gereksinimi, veri setini uygun bir olasılık dağılımı ile modellemek ve bu modelin parametrelerini tahmin etmektir.

### II. KULLANILAN YÖNTEMLER

Bu problemde gözetimli öğrenme tekniklerinden biri olan doğrusal regresyon ile gözetimli öğrenme yöntemi kullanılmıştır.

#### A. GÖZETİMLİ ÖĞRENME

Gözetimli öğrenme, bir makinenin verileri öğrenmek ve tahminlerde bulunmak için etiketlenmiş veri kullanarak eğitildiği bir makine öğrenme alt dalıdır. Bu süreçte, veri noktaları girdi (veri) ve çıktı (etiket) olarak çiftler halinde gelir. Algoritma, bu etiketli veri kümesini kullanarak girdileri çıktılarla ilişkilendiren bir model oluşturur. Daha sonra bu model, yeni, etiketlenmemiş verilere dayalı tahminler yapabilir.



Şekil 1. Doğrusal Regresyon Biçim Örneği

#### B. DOĞRUSAL REGRESYON İLE GÖZETİMLİ ÖĞRENME

Doğrusal regresyon, istatistiksel bir yöntemdir ve değişkenler arasındaki ilişkiyi modellemek için kullanılır. Bu yöntem, farklı veri noktalarını analiz ederek bir doğru çizer. Katsayılar, bağımsız değişkenler ile bağımlı değişken arasındaki ilişkinin büyüklüğünü ve yönünü temsil eder. Doğrusal regresyon denklemi, bağımsız değişkenlerin değerleri verildiğinde bağımlı değişkenin tahmini değerini hesaplamak için kullanılır. (Şekil 1)

### III. PROBLEM ANALİZİ VE ÇÖZÜMÜ

#### A. Kütüphane Analizi

Problem analizinde ilk olarak belirtilen kütüphaneler ve kütüphanelerdeki fonksiyon yapıları incelenmiştir.

#### B. Veri Yükleme

Samples.txt dosyasındaki verileri modele uygun hale getirebilmek için NumPy kütüphanesini kullanarak bir NumPy dizisi döndüren bir işlem gerçekleştirildi. Bu işlemin amacı, dosyadaki verileri yüklemek ve verileri modellemeye uygun bir formata dönüştürmektir.

#### C. Parametrelerin Bulunması

Verileri kullanılabilir hale getirdikten sonra parametrelerin analizini yapmak için kullanılan denklemler Şekil 2 Şekil 3 ve Şekil 4'te gösterilmiştir.

Sınıflandırmada, verilen bir girdiye karşılık Boolean bir çıktı (evet/hayır) elde edilirken, çıktı nümerik bir değerse, öğrenmek istediğimiz şey bir sınıf değil, nümerik bir fonksiyondur; bu fonksiyon bilinmemekte ancak bu fonksiyondan çekilen örneklerden oluşan bir eğitim kümesine sahip olunmaktadır.

$$\mathcal{X} = \{\mathbf{x}^t, r^t\}_{t=1}^N$$

Şekil 2.

Eğer gürültü yoksa, görev interpolasyondur ve bu noktaların üzerinden geçen ve Şekil 3'te bulunan şartı sağlayan fonksiyon  $f(\mathbf{x})$ 'i bulmak istiyoruz.

$$r^t = f(\mathbf{x}^t)$$

$$r^t = f(\mathbf{x}^t) + \epsilon$$

$$r^t = f^*(\mathbf{x}^t, \mathbf{z}^t)$$

$$E(g|\mathcal{X}) = \frac{1}{N} \sum_{t=1}^N [r^t - g(\mathbf{x}^t)]^2$$

$$g(\mathbf{x}) = w_1x_1 + \dots + w_dx_d + w_0 = \sum_{j=1}^d w_jx_j + w_0$$

Şekil 3.

# ELM472 Makine Öğrenmesinin Temelleri

'r' ve 'g(x)' nümerik değerler oldukları için değerleri üzerinde bir sıralama tanımlayabiliriz ve bu değerler arasında bir mesafe tanımlayabiliriz; bu, sınıflandırmada kullanılan eşit/değil eşit'ten daha fazla bilgi sağlar. Amacımız, deneysel hatayı en aza indiren 'g(·)' fonksiyonunu bulmaktır ve bu fonksiyon için belirli bir varsayım kümesi ve parametre seti kullanmalıyız.

$$g(x) = w_1x + w_0$$

Şekil 4.

$$E(w_1, w_0 | \mathcal{X}) = \frac{1}{N} \sum_{t=1}^N [r^t - (w_1x^t + w_0)]^2$$

Şekil 5.

Şekil 4 ve 5'te gösterilen  $w_1$  ve  $w_0$  değerleri verilerden öğrenilecek parametrelerdir.  $w_1$  ve  $w_0$  değerleri en aza indirilmelidir.

Minimum noktası, E'nin  $w_1$  ve  $w_0$ 'a göre kısmi türevleri alınarak, bunları 0'a eşitleyerek ve iki bilinmeyeni çözerek Şekil 4'te gösterildiği gibi hesaplanabilir.

$$w_1 = \frac{\sum_t x^t r^t - \bar{x} \bar{r} N}{\sum_t (x^t)^2 - N \bar{x}^2}$$
$$w_0 = \bar{r} - w_1 \bar{x}$$

Şekil 6.

Bu denklemler kullanılarak hesaplamalar kod içerisinde yapılmıştır.

## D. Verilerin Görselleştirilmesi

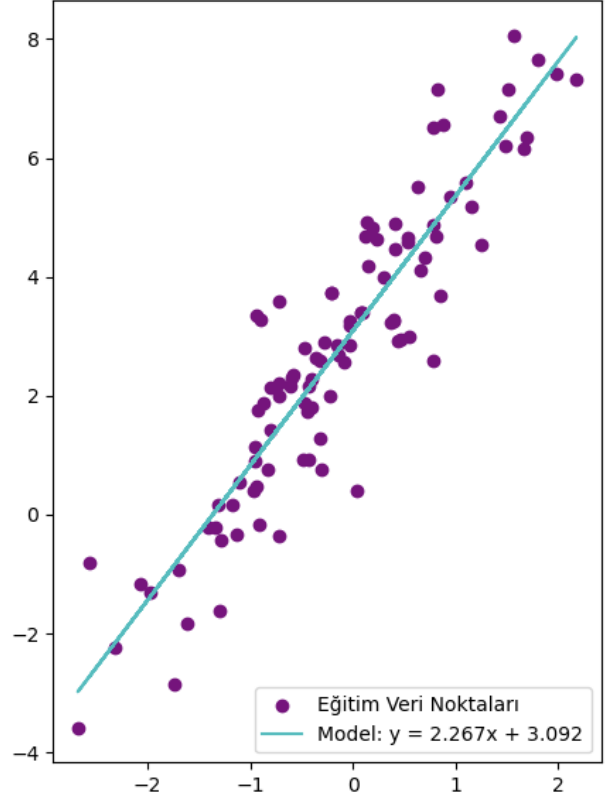
Dağılım grafiği oluşturuldu. Bu grafik sayesinde eğitim verileri temsil edildi. Lineer regresyon modeli, elde edilen  $w_1$  ve  $w_0$  parametreleri kullanılarak grafik üzerinde belirtildi.

## E. Fonksiyonların Kullanılması

Ana fonksiyon yapısında; verilerin okunarak sütunlar halinde farklı değişkenlere atama, lineer regresyon model katsayılarını hesaplama ve elde edilen sonuçlar ile lineer regresyon modelini grafikte görselleştirme işlemleri sırasıyla gerçekleştirildi.

## IV. SONUÇ VE YORUMLAR

Eğitim Veri Noktaları ve Lineer Regresyon Modeli



Şekil 5.

Lineer regresyon yöntemi, veri noktaları arasındaki ilişkiyi modellemek için  $g(x|\theta)$  şeklinde temsil edilebilen bir model kullanır. Modelde  $g(\cdot)$ , veri noktaları arasındaki lineer ilişkiyi temsil ederken,  $x$  girdiyi ve  $\theta$  da modelin eğim ve kesiş parametrelerini temsil eder. Bu probleme çözüm olarak; verilen  $x$  girdilerini, ilgili istenen  $r^t$  çıktılarına faydalı bir yaklaşım oluşturacak parametreleri kullanarak modelleme yapıldı.

Şekil 5'te görüleceği üzere, veri noktaları arasındaki ilişki modellendi. Kod yardımı ile veri noktalarını temsil eden dağılım grafiği ve bu noktalar arasında en uygun doğruyu (lineer modeli) gösterildi. Modele uygun, denklemlere dayanarak en uygun parametreler ( $w_1$  ve  $w_0$ ) hesaplandı. Bu parametreler, veri noktalarının dağılımıyla en uyumlu doğruyu elde etmek için kullanıldı. Bu doğrunun çizilme amacı değişkenler arasındaki ilişkiyi modellemek içindir.

## KAYNAKÇA

- [1] E. Alpaydin, Introduction to Machine Learning, 3. bs. Cambridge, MA, USA: MIT Press, 2014