

ELM472 – Makine Öğrenmesinin Temelleri

Ödev 1

Son teslim tarihi: 11.09.2022 – 13:00

Alican Bayındır
a.bayindir2020@gtu.edu.tr
Elektronik Mühendisliği Bölümü, GTÜ, Kocaeli, Türkiye

I. GİRİŞ

Regresyon analizi, bir bağımlı değişken (genellikle “sonuç” veya “yanıt” değişkeni veya makine öğrenimi jargonunda “etiket” olarak anılır) ile bir veya daha fazla arasındaki ilişkileri belirlemek için istatistiksel modellemede kullanılan bir grup istatistiksel prosedürdür. Bağımsız değişkenler (genellikle “öngörücüler”, “ortak değişkenler” olarak anılır.) Doğrusal regresyonda, en tipik regresyon analizi türü belirli bir matematiksel kriter açısından verilerle en yakından eşleşen doğru (veya daha karmaşık bir doğrusal kombinasyon) bulunur. Sıradan en küçük kareler yaklaşımını kullanırken, örneğin, gerçek veriler ile bu doğru arasındaki kare farklarının toplamını en aza indiren belirli bir çizgi (veya hiperdüzlem - hyperplane) Bu, bağımsız değişkenler kesin matematiksel nedenlerle belirli bir dizi değer alındığında, araştırmacının bağımlı değişkenin koşullu beklentisini (veya popülasyon ortalama değerini) tahmin etmesini sağlar. [1]

II. UYGULAMA

A. Regresyon doğrusunun bulunması

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

$$w_0 = \bar{r} - w_1 \bar{x}$$

Formül 1 Ödev'de bulunması istenen parametreleri bulmak için kullanılan formüller.[2]

Formül 1'deki formül ders kitabından alınmıştır. Formülün kullanımı konusunda isimlendirme -notasyon-kod kısmında farklıdır. Ancak mantık olarak aynı mantık kullanılmıştır. [3]

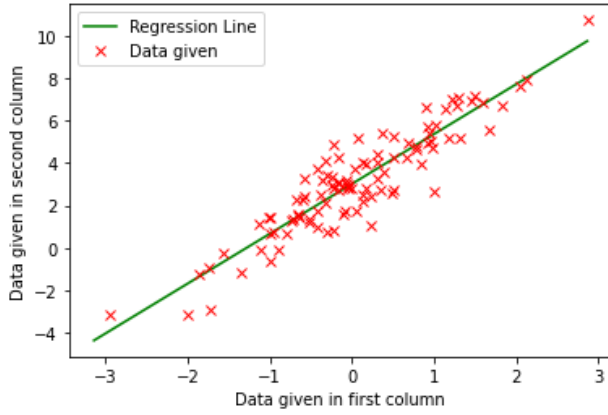
B. Python kodunun yazılması

Gerekli araştırma ve konu incelenmesi yapıldıktan sonra Python kodu yazılmış ve ekte iletilmiştir. Formülün koda işlenmesi sırasında ilk olarak izin verilen 2 kütüphane ödevde bize teslim edilen veri dosyası ile birlikte içeri aktarılmıştır. (Numpy ve matplotlib kütüphaneleri). Daha sonra içe aktarılan veri dosyası satırlarına göre parçalara ayrılmış olup formül gereği ortalama değerleri bulunmuştur. Sonrasında Formül 1'de verilen formüldeki işlemler numerator ve denominator isimleri altında 2'ye ayrılmıştır ve for döngüsü içerisinde bunların değerleri hesaplanmıştır. Bu işlemler neticesinde w_1 değeri ~ 2.348 , w_0 değeri ~ 3.009 bulunmuştur. -Ödevin a şıkkına ithafen-

Bütün bunlar hesaplandıktan sonra ödevin B şıkkındaki isteri gerçekleştirmek için; veriler önce ham hali x işareti ile gösterildi ve hesaplanan regresyon doğrusu bu verilerin üzerine çizdirildi. Çizdirilen grafik aşağıda, Şekil 1 adı altında paylaşılmıştır.

III. SONUÇ

Ödevde bize teslim edilen veriler başarılı bir şekilde analiz edilmiş olup, regresyon çizgisi hesaplanmıştır. Bu yöntem bağımlı ve bağımsız değişkenler arasındaki ilişkiyi belirlemeye yönelik istatistiksel bir analiz yöntemidir (Regresyon yöntemi). Bağımsız faktörler değiştiğinde bağımlı değişkenin nasıl değiştiğine daha fazla odaklanır. Ayrıca gelecekte farklı faktörlerin nasıl etkileyeceğini modellemeye yardımcı olur. Şekil 1'de yeşil renk ile çizdirilen regresyon çizgisine bakıldığında elimizde tam olarak ilgili noktadaki değere karşılık bir ölçüm veya veri olmasa da çizgi üzerinde takip edilerek istenen noktada tahmini bir şekilde ilgili x değerinin ortalama olarak hangi y değerine denk geleceği görülebilir. Bu yöntem pazarlama, üretim, ilaç, finans sektörleri hakkında öngörü – tahminde bulunmak için kullanılır.[1][4]



Şekil 1 Regresyon çizgisi ve veri noktalarının grafik üzerinde çizdirilmiş hali.

IV. YAZARLAR

- 1- Alican Bayındır – Gebze Teknik Üniversitesi, Elektronik Mühendisliği, Lisans Öğrencisi, Kocaeli, Gebze.

V. TEŞEKKÜR

Projeyi yaparken unuttuğum Python bilgisini tazelemek amacıyla geçtiğimiz dönem verdiğim ELM368 dersinde yaptığım projeleri inceledim. Bütün bu projeleri yapmama ve Python bilgisi elde etmeme sebep olan GTU Elektronik Mühendisliği Dr. Öğr. Üyesi Köksal HOCAOĞLU'na teşekkürlerimi sunuyorum.

VI. KAYNAKÇA

- [1] Ahmet Güneş, Dr. Öğr. Üyesi, ELM 472 - Makine Öğrenmesi Temelleri dersi, Gebze Teknik Üniversitesi.
- [2] E. Alpaydin, Introduction to Machine Learning, 3. bs. Cambridge, MA, USA: MIT Press, 2014.
- [3] Mirko Stojiljković. (2022, May 16). Linear Regression in Python. Real Python. <https://realpython.com/linear-regression-in-python/>
- [4] Andrew NG, Dr. , CS 229 - Makine öğrenmesi, Harvard Üniversitesi