

Galatasaray Üniversitesi
Bilgisayar Mühendisliği Bölümü
INF102 - Programlamaya Giriş

Ödev 06

Dikkat! Programı derleme hatası tetikleyenler doğrudan sıfır alacaktır. Ödevinizi göndermeden önce çalıştığını mutlaka kontrol edin, eksik olsa da düzgün şekilde çalışan bir kod gönderin. Gerekirse hatalı kısımları yorum satırına alıp, bana not yazın.

Katsayı Tahmini (100 puan)

Bu ödevinizde, şimdiye kadar derste gördüğümüz tüm konuları kullanarak basit bir makine öğrenmesi simülasyonu yapacaksınız. Makine öğrenmesinde en sık çalışılan konulardan biri tahminlemedir. Mesela elimizde bazı veriler olduğunu varsayalım, bu verileri modelleyip onlara en düzgün şekilde uyacak modeli bulmak isteriz ki, elimizde olmayan bir veri ile karşılaştığımız zaman onu nereye nasıl yerleştireceğimizi bilelim.

Aslında bu, günlük hayatta çok sık karşılaştığımız şeylerden biridir. Hava durumundan tutun da aylık ortalamama harcamanızı, herhangi bir maçın skorunu, sene sonu harf notunuzu, dövizin ne kadar yükselip alçalacağını, akşam okuldan çıkınca eve yaklaşık kaç dakikada döneceğinizi hesaplamaya kadar hayatınızın bir çok alanında deneyimlediğiniz bir şey.

Bu ödevde bu deneyimi bir denklemin katsayılarını tahmin etmek için kullanacaksınız. Bol bol açıklama yaptığım ve bir takım yönlendirmelerde bulunduğum yorum satırlı `odev06.c` kaynak kod dosyasını Moodle'dan indirdiğiniz bu .zip dosyası içerisinde bulabilirsiniz.

Sizinle paylaştığım `data.csv` isimli dosya içerisinde 250 adet x ve bunlara karşılık gelen y koordinat bilgileri bulunuyor. Bu y verileri aşağıdaki denklemden geliyor:

$$y = ax^4 + bx^3 + cx^2 + dx + e \quad x, y \in \mathbb{R}; a, b, c, d, e \in [1, 2, 3, \dots, 9] \quad (1)$$

Ve sizden 1 no'lu denklemin değerleri 1 ile 9 arasındaki tamsayılardan oluşan $\{a, b, c, d, e\}$ katsayılarını tahmin etmeniz isteniyor. Yaptığımız tahminin ne kadar iyi bir tahmin olduğunu değerlendirebilmeniz için de bir hata ölçeği kullanmanız gerekiyor, bu amaçla denklemini aşağıda verilen ortalama kare hata (*Mean Squared Error*) değerini kullanacaksınız:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2)$$

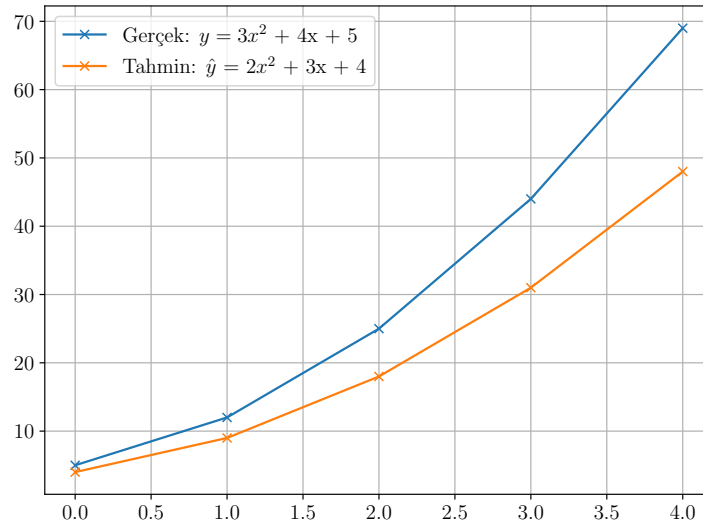
2 no'lu denklemde n örnek sayısını, y_i i . örneğin gerçek değerini, \hat{y}_i ise i . örneğin tahmini değerini temsil ediyor.

Yukarıdaki bilgileri bir örnek üzerinden tekrarlayalım. Aşağıda denklemleri verilen iki modeli karşılaştıralım, y aslında katsayılarını tahmin etmeye çalıştığımız gerçek model, \hat{y} ise ilk tahminimiz olsun:

$$y = 3x^2 + 4x + 5 \quad (3)$$

$$\hat{y} = 2x^2 + 3x + 4 \quad (4)$$

$x = \{0, 1, 2, 3, 4\}$ olduğunda $y = \{5, 12, 25, 44, 69\}$ ve $\hat{y} = \{4, 9, 18, 31, 48\}$ değerlerini alıyor (denklemlerin grafikleri Şekil 1'de görselleştirilmiştir).



Şekil 1: Gerçek fonksiyon ve tahmin

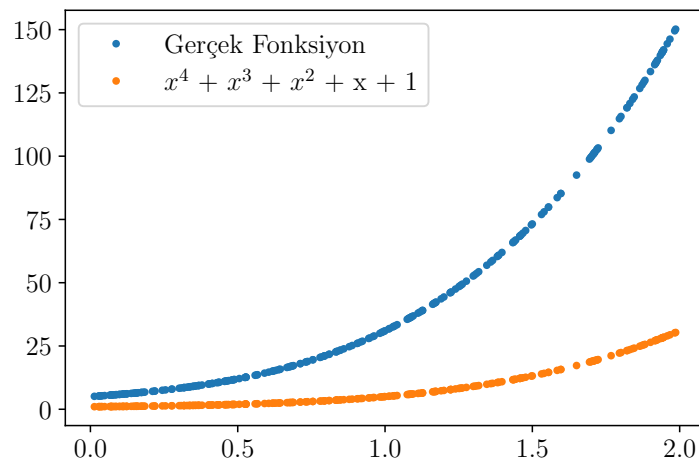
Tahminimizdeki hata payını hesapladığımızda da, MSE 5 no'lu denklemdaki değeri alıyor:

$$MSE = \frac{1}{5} [(5 - 4)^2 + (12 - 9)^2 + (25 - 18)^2 + (44 - 31)^2 + (69 - 48)^2] = 133.8 \quad (5)$$

Burada amacımız bize minimum hatayı veren katsayı setini bulmak, ödevinizde sizden aynı işlemi 4. dereceden bir denklem için yapmanız isteniyor ve size verilen 5 farklı katsayı setinden hangisinin daha iyi sonuç verdiğini hesaplayıp açıklamamız bekleniyor.

BONUS: İsteyenler bonus olarak modelin katsayılarını bulan kodu da ekleyip gönderebilirler.

Not: Kaynak kod dosyasındaki örnekte, yaptığımız ilk tahmin ($\hat{y} = x^4 + x^3 + x^2 + x + 1$) ile katsayılarını tahmin etmeye çalıştığımız y denklemini karşılaştırmalı görmek isterseniz, Şekil 2'yi inceleyebilirsiniz.



Şekil 2: Gerçek model ve odev06.c dosyasındaki ilk örnekte kullanılan tahmini model