

Multiple Linear Regression

- Basit doğrusal regresyon modelinin iki veya daha fazla bağımsız değişkene genişletilmesi.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon$$

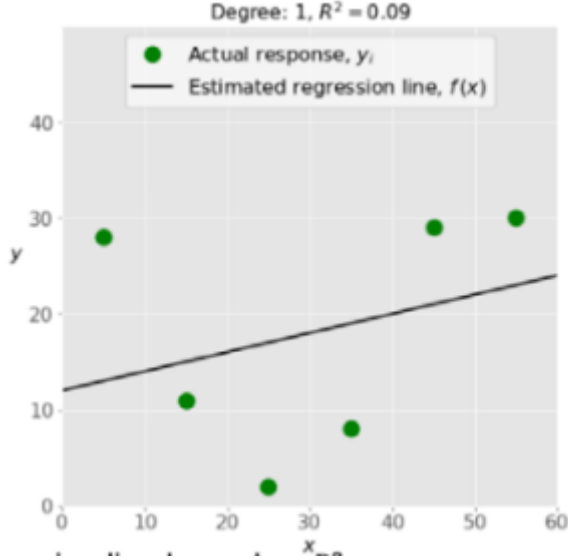
$$\text{Expression} = \text{Baseline} + \text{Age} + \text{Tissue} + \text{Sex} + \text{Error}$$

- Kısmi Regresyon Katsayıları: $\beta_i \equiv$ bağımsız değişkeni 1 birim arttırırken bağımlı değişken üzerindeki etki, diğer tüm yordayıcıları sabit tutar.

Underfitting and Overfitting

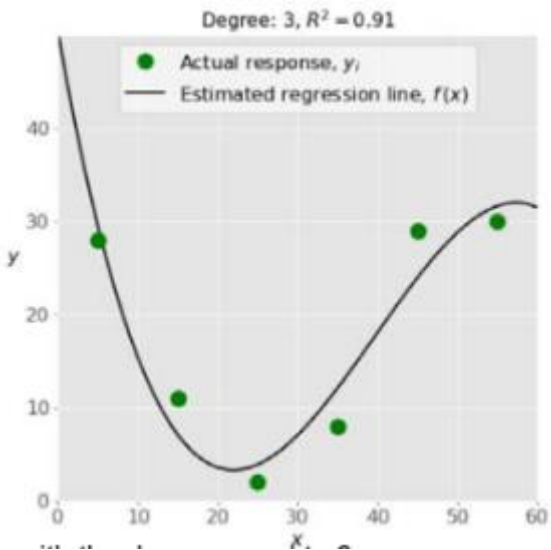
- Yetersiz yerleştirme, bir model genellikle kendi sadeliğinin bir sonucu olarak veriler arasındaki bağımlılıkları doğru bir şekilde yakalayamadığında ortaya çıkar.
- Yeni verilerle uygulandığında genellikle bilinen veriler ve düşük genelleme özellikleri ile düşük bir R2 verir.
- Aşırı sığdırma, bir model hem veriler arasındaki bağımlılıkları hem de rastgele dalgalanmaları öğrendiğinde gerçekleşir.
- Diğer bir deyişle, bir model mevcut verileri çok iyi öğrenir.
- Pek çok özelliği veya terimi olan karmaşık modeller, genellikle aşırı sığmaya eğilimlidir.
- Bilinen verilere uygulandığında, bu modeller genellikle yüksek R2 verir.
- Ancak, genellikle iyi bir genelleme yapmazlar ve yeni verilerle kullanıldığında önemli ölçüde daha düşük R2'ye sahiptirler.

Linear Regression: Underfitting Example



- Bu doğrusal regresyon hattının R^2 değeri düşüktür.
- Düz çizgi, x 25'ten sıfıra doğru ilerledikçe gerçek yanıtın arttığı gerçeğini hesaba katmaz.
- Bu muhtemelen yetersiz uyumun bir örneğidir.

Linear Regression: Overfitting Example



- Derecesi 3'e eşit olan polinom regresyon.
- R^2 değeri önceki durumlardan daha yüksektir.

- Bu model, bilinen verilerle öncekilerden daha iyi davranır.
- Bununla birlikte, gerçek veriler bunu göstermese de, özellikle hattın azalmaya başladığı 60'a yakın giriş değerleri için bazı fazla sığdırma belirtileri gösterir.

Kategorik Bağımsız Değişkenler

- Nitel değişkenler kukla değişkenler aracılığıyla regresyon çerçevesine kolayca dahil edilir
- Basit örnek: seks 0/1 olarak kodlanabilir
- Kategorik değişkenim üç seviye içeriyorsa:



- Önceki kodlama, doğrusallık ile sonuçlanır
- Çözüm, bir dizi sahte değişken oluşturmaktır. Genel olarak, k seviyeleri için k-1 kukla değişkenlerine ihtiyacınız vardır

$$x_1 = \begin{cases} 1 & \text{if AA} \\ 0 & \text{otherwise} \end{cases}$$

$$x_2 = \begin{cases} 1 & \text{if AG} \\ 0 & \text{otherwise} \end{cases}$$

	x_1	x_2
AA	1	0
AG	0	1
GG	0	0