

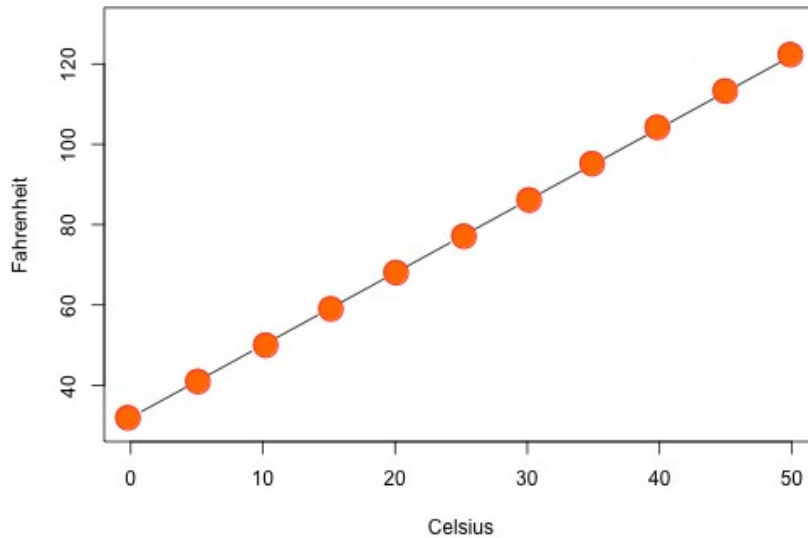
3.2 Basit Lineer Regresyon

Başta istatistik olmak üzere neredeyse tüm bilim dallarında ve mühendislikte değişkenler arasındaki ilişki ölçülmek istenmektedir. Değişkenler iki veya daha çok sayıda olabilir. Türkçe olarak bağılanım veya daha yaygın olarak kullandığımız regresyon, iki ya da daha çok değişken arasındaki ilişkiyi ölçmek için kullanılan bir analiz metodudur. Tek bir değişken kullanılarak regresyon analizi gerçekleştiriliyorsa buna tek değişkenli regresyon, birden fazla değişken kullanılıyorsa da çok değişkenli regresyon analizi denir. Regresyonda, değişkenlerden biri bağımlı diğerleri bağımsız değişken olmalıdır.

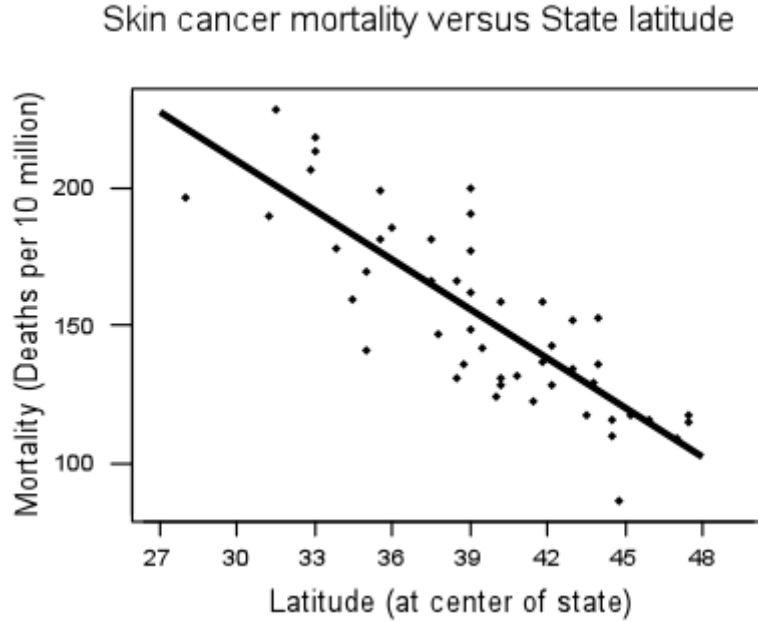
3.2.1 İlişki Türleri

Değişkenler arasındaki ilişki türlerini **deterministik (veya fonksiyonel)** ve **istatistiksel ilişki** olarak ikiye ayırabiliriz. Deterministik ilişkilerin her birinde, değişkenler için yazılan denklem iki değişken arasındaki ilişkiyi tam olarak tanımlamaktadır. Deterministik ilişkilere örnek olarak temel bilimlerde kullanılan formülleri verebiliriz. Örneğin celsius cinsinden ölçülen hava sıcaklığının fahrenheit olarak karşılığı bellidir. Aşağıdaki formülü kullanarak bir bakıma dönüşüm yaptığımızda sıcaklığın değerini fahrenheit cinsinden tam olarak biliriz. Deterministik ilişkilerin grafiklerini çizecek olsaydık, celsius değerlerine karşılık gelen fahrenheit değerlerinde olduğu gibi tüm (x, y) veri noktalarının doğrudan bir çizgiye düştüğünü görürdük.

$$F = \left(C * \frac{9}{5} \right) + 32$$



Regresyon, deęişkenler arasındaki ilişkinin deterministik ilişkilerdeki gibi mükemmel olmadığı istatistiksel ilişkilerle ilgilenir. İstatistiksel ilişkide deęişkenlerin grafiğine baktığımızda eğilim sergilerken aynı zamanda bir dağılım da sergiler. İstatistiksel ilişkide deęişkenlerin dağılım grafięi aşığıdaki gibi olabilir.



3.2.2 İlişkinin Gücü ve Yönü

İki rassal deęişken arasındaki doğrusal ilişkinin yönünü ve gücünü **korelasyon** belirtir. İki deęişken arasındaki ilişki miktarı basit korelasyon denen korelasyon teknikleri ile hesaplanır. Bir deęişkenin iki veya daha fazla deęişken ile olan ilişkisi çoklu korelasyondur. Bu deęişkenlerden birinin sabitlenerek dięer deęişkenler ile olan ilişkisi kısmi korelasyon teknikleri ile hesaplanır.

Deęişkenler arasındaki ilişki, korelasyon katsayısı ile hesaplanmaktadır. Birçok korelasyon katsayısı bulunmaktadır. Bunlardan hangisinin kullanılacağına deęişkenlerin özelliklerine göre karar verilir. Deęişkenlerin sürekli veya kesikli olması, doğrusal olup olmaması gibi özellikler hangi korelasyon katsayısının kullanılacağını etkiler.

Basit Korelasyon, Pearson Momentler Çarpımı Korelasyon Katsayısı

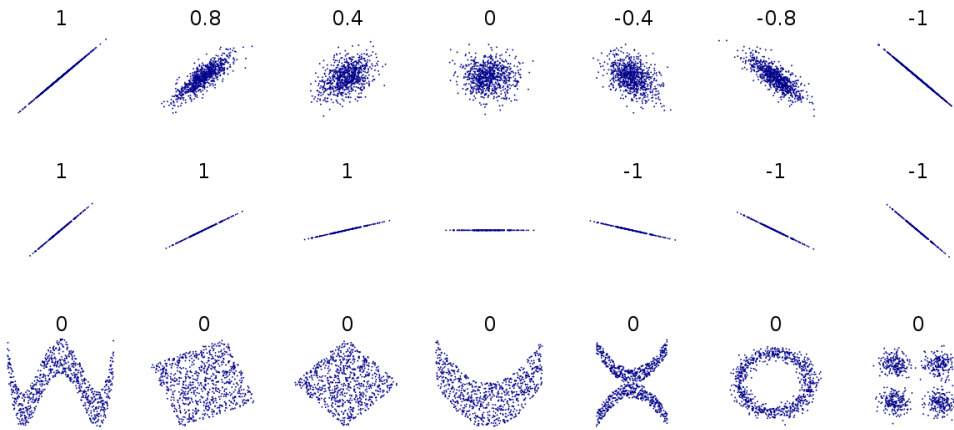
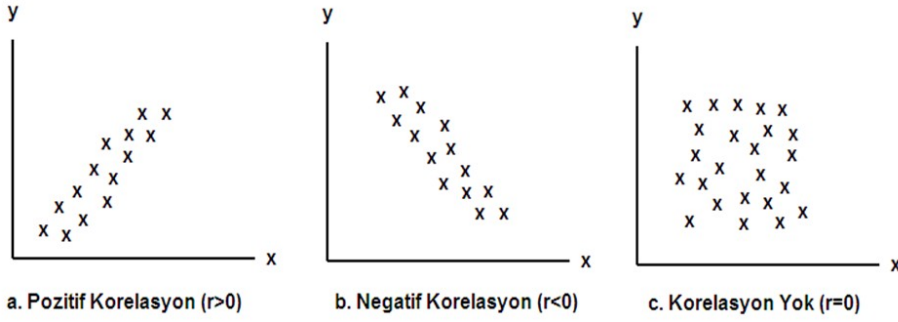
Basit korelasyon, iki deęişken arasındaki ilişkiyi belirtir. Pearson momentler çarpımı korelasyon katsayısı, iki deęişkenin kovaryansının, yine bu deęişkenlerin standart sapmalarının çarpımına bölünmesiyle elde edilir. Korelasyon analizi sonucunda, deęişkenler arasında doğrusal ilişki olup olmadığı ve varsa aralarındaki ilişkinin derecesi Pearson momentler çarpımı korelasyon katsayısı ile hesaplanabilir.

Pearson momentler çarpımı:

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

İlişkileri de aşağıdaki gibi yorumlayabiliriz.

Şekil 1. Farklı Korelasyon Durumları



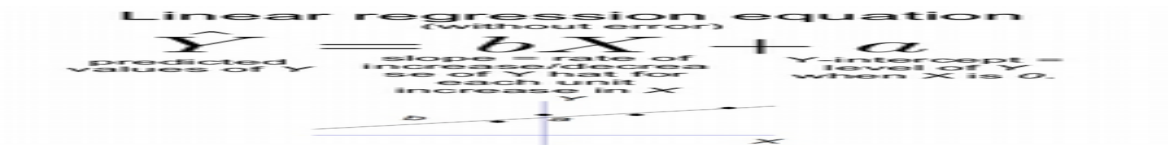
Pearson Momentler Çarpımı Korelasyon katsayısı -1 ile +1 arasındaki değerleri alabilir.
Pearson Momentler Çarpımı Korelasyon Katsayısı (r),

- $r = -1$ ise tam negatif doğrusal ilişki
- $r = +1$ ise tam pozitif doğrusal ilişki
 - 0.01 - 0.29 düşük düzeyde ilişki
 - 0.30 - 0.70 orta düzeyde ilişki
 - 0.71 - 0.99 yüksek düzeyde ilişki
- $r = 0$ ise aralarında ilişki yoktur

diyebiliriz.

Regresyonun tanımını biraz daha açıklayacak olursak aralarında istatistiksel ilişki olan iki veya daha çok değişkenden birinin bağımlı, diğerlerinin bağımsız değişkenler olarak ayrılan ve değişkenler arasındaki ilişkinin matematiksel bir eşitlik ile açıklanması sürecidir diyebiliriz.

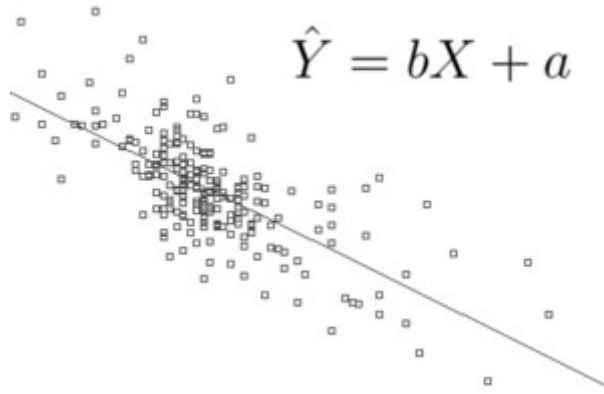
Basit Lineer Regresyonun denklemi şu şekildedir:



- X olarak adlandırılan değişken, öngörücü, açıklayıcı veya bağımsız değişken
- Y olarak adlandırılan diğer değişken, yanıt, sonuç, açıklanan veya bağımlı değişken

olarak kabul edilir.

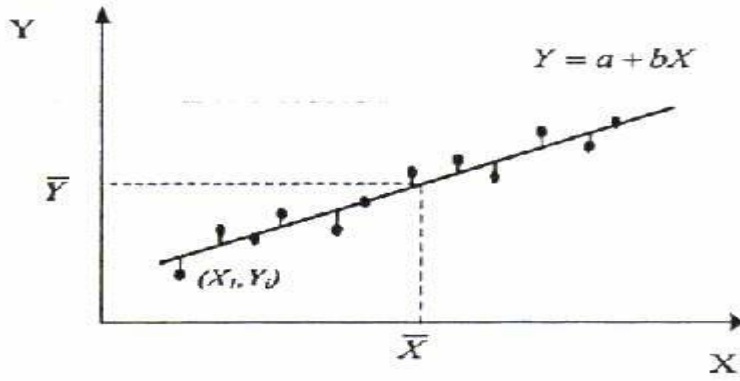
Aralarında negatif korelasyon bulunan iki değişken ve basit doğrusal regresyon eğrisi aşağıdaki gibi olacaktır.



3.2.3 En Küçük Kareler Yöntemi

En küçük kareler yöntemi, regresyon problemleri ele alındığında veri setine en iyi uyan fonksiyonu yazabilmek için çok kullanılan bir yöntemdir. En küçük kareler yöntemi, regresyon denklemi ile hesaplanan tahmini veri noktaları ile veri setinde verilen her bir veri noktası arasındaki uzaklıkların kareleri toplamını minimum yapan a ve b katsayılarını bulma işlemidir. Bu sayede veri seti ile gelen tüm veri noktalarına mümkün olduğu kadar yakın geçecek bir fonksiyon eğrisi bulunmuş olur.

Enküçük Kareler Yöntemi:



İlk minik makine öğrenmesi uygulamasını gerçekleştirirken aslında regresyon eğrisini en küçük kareler yöntemi ile bulmuştuk. En küçük kareler yönteminin matematiksel ispatlarına girmeden formulleri verecek olursak:

Regression Formula:

$$Y = a + bX$$

where slope of trend line is calculated as:

$$b_1 = \frac{\sum (x - \bar{x}) * (y - \bar{y})}{\sum (x - \bar{x})^2}$$

and the intercept is computed as:

$$b_0 = y - (b_1 * X)$$

$$\beta_i = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \bar{\beta}_1 \bar{x}$$

$$m = \frac{n \sum (xy) - \sum x \sum y}{n \sum (x^2) - (\sum x)^2} \quad b = \frac{\sum y - m \sum x}{n}$$