Regresyon Ve Korelasyon

Dr. Cahit Karakuş

2021 - İsatnbul

İçindekiler

1. Giriş

Makine öğrenmesinde denetemli öğrenme algoritması olan regresyon modelinin yapısı çok boyutlu doğrusal olmayan en küçük kareler yöntemi için temel oluşturan destek değerlerinin düzeltilmesine dayanır.

Standart eğriler:

- *Düz çizgi*
- *Doğrusal denklemler, y=at+b*
- *İkinci dereceden eğriler*
- *Üçüncü dereceden eğriler*
- *Çember*
- *Elips*
- *Logaritmik eğriler*
- *Üstel eğriler*
- *Hiperbolik eğriler*
- *Trigonometrik eğriler*

2. Korelasyon Analizi

Korelasyon analizi, bir fonksiyonun değişkenleri arasındaki ilişkinin yönünü, derecesini ve önemini ortaya koyan istatistiksel yöntemdir. Değişkenler arasındaki ilişkinin yönünü ve derecesini belirten katsayıya korelasyon katsayısı denir.

Korelasyon katsayısı küçük r harfi ile gösterilir ve r değeri -1 ile +1 arasında değerler alır. Eğer r değeri -1'e yakın değerler alıyor ise değişkenler arasında negatif yönde, +1'e yakın değerler alıyor ise pozitif yönde bir ilişki olduğu belirlenir. Eğer r değeri sıfıra yakın değerler alıyor ise iki değişken arasında bir ilişki olmadığı sonucuna varılır.

Korelasyon katsayısı negatif ise iki değişken arasında ters ilişki vardır, yani "değişkenlerden biri artarken diğeri azalmaktadır" denir. Korelasyon katsayısı pozitif ise "değişkenlerden biri artarken diğeride artmaktadır" yorumu yapılır.

Çok sayıda korelasyon analizi mevcuttur. Ancak en yaygın kullanılan korelasyon analizleri;

- Pearson çarpım moment korelasyon katsayısı
- Spearman'ın sıralama korelasyon katsayısı

Verilerin normal dağılıma sahip olması durumunda Pearson korelasyon katsayısı, verilerin normal dağılmadığı durumda ise Spearman Rank korelasyon katsayısı tercih edilir. Bir korelasyon katsayısının yorumlanabilmesi için p değerinin 0.05 den daha küçük olması gerekir.

Pearson Çarpım Moment Korelasyon Katsayısı:

$$r = \frac{n(\sum x_i y_i) - (\sum x_i)(\sum y_i)}{\sqrt{n(\sum x_i^2) - (\sum x_i)^2}\sqrt{n(\sum y_i^2) - (\sum y_i)^2}}$$

Korelasyon katsayısı (r) nın yorumu;

- r<0.2 ise çok zayıf ilişki yada korelasyon yok
- 0.2-0.4 arasında ise zayıf korelasyon
- 0.4-0.6 arasında ise orta şiddette korelasyon
- 0.6-0.8 arasında ise yüksek korelasyon
- 0.8> ise çok yüksek korelasyon olduğu yorumu yapılır.

Örnek:

SUBJECT	AGE X	GLUCOSE LEVEL Y	XY	X^2	Y^2
1	43	99	4257	1849	9801
2	21	65	1365	441	4225
3	25	79	1975	625	6241
4	42	75	3150	1764	5625
5	57	87	4959	3249	7569
6	59	81	4779	3481	6561
Σ	247	486	20485	11409	40022

The answer is: r=**2868 / 5413.27 = 0.529809**

From our table:

- Σx = 247
- Σy = 486
- Σxy = 20,485
- Σx^2 = 11,409
- Σy^2 = 40,022
- n is the sample size, n = 6

The correlation coefficient = 6(20,485) – (247 × 486) / [√[[6(11,409) – (247^2)] × [6(40,022) – 486^2]]]

= 0.5298

The range of the correlation coefficient is from -1 to 1. Our result is 0.5298 or 52.98%, which means the variables have a moderate positive correlation.

Örnek:
Σx = 60, Σy = 60, Σxy = 400, Σx^2 = 400, Σy^2= 400, n = 10

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

r=1, çok yüksek korelasyon olduğu yorumu yapılır

Örnek:
Aşağıdaki veriler için korelasyon katsayısı r'yi hesaplayın.

x	y	xy	x^2	y^2
1	-3	-3	1	9
2	-1	-2	4	1
3	0	0	9	0
4	1	4	16	1
5	2	10	25	4
$\Sigma x = 15$	$\Sigma y = -1$	$\Sigma xy = 9$	$\Sigma x^2 = 55$	$\Sigma y^2 = 15$

$$r = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{n\Sigma x^2 - (\Sigma x)^2}\sqrt{n\Sigma y^2 - (\Sigma y)^2}} = \frac{5(9) - (15)(-1)}{\sqrt{5(55) - 15^2}\sqrt{5(15) - (-1)^2}}$$

$$= \frac{60}{\sqrt{50}\sqrt{74}} \approx 0.986$$

X ve y arasında güçlü bir pozitif doğrusal korelasyon vardır.

Spearman'ın sıralama korelasyon katsayısı:
İstatistik bilim dalında, Spearman'ın sıralama korelasyon katsayısı ismi verilen istatistiksel ölçüyü ilk ortaya atan Amerikan istatistikçi Charles Spearman'a atfen adlandırılmıştır. Matematik notasyon olarak çok defa eski Yunan harfi ρ (rho okunur) ile belirtilir.

Parametrik olmayan istatistik ölçüsüdür ve iki değişken arasındaki bağımlılık, yani korelasyon, ölçüsü olarak bulunup kullanılır. Bu demektir ki Spearman'in ρ katsayısı iki değişken için çokluluklar dağılımı hakkında hiçbir varsayım yapmayarak, bu iki değişken arasında bulunan bağlantının herhangi bir monotonik fonksiyon ile ne kadar iyi betimlenebilineceğini değerlendirmek amaçlı incelemedir.

Prensip olarak Spearman'ın sıralama korelasyon katsayısı ρ, Pearson çarpım-moment korelasyon katsayısının özel bir halidir. ρ değerinin hesaplanması için iki değişken (Y ve X) içinde örneklem verilerinin sıralama düzeninde olmaları gereklidir. Genel olarak, örneklem verileri için bu koşul uygun değildir ve veriler sıralama düzeni halinde olmadan oransal ölçekli veya aralıksal ölçekli veya sırasal ölçekli olarak bulunur ve bu halde bir dönüşümle sıralama düzeni haline sokulurlar. Böylece ρ formulü için sıralama düzenli xi ve yi örneklem verileri kullanılır.

Sonra iki değişken için karşılıklı veri elemanları (xi ve yi)'nin sıra numaraları arasındaki fark di, i=1,...n olarak bulunur. Bu tüm karşılıklı veriler (i=1...n) için uygulanır. Eğer sıra numaraları arasında hiç beraberlik yoksa, ρ değerini bulmak için şu formül kullanılır:

$$\rho = 1 - \frac{6\sum_{i=1}^{n} d_i}{n(n^2-1)}$$

Burada
di=xi-yi: i elamanı Xi ile Yi sıra numaraları arasındaki fark;
n : iki değişkenli örneklemde toplam gözlem sayısıdır.

Örnek:

Aşağıdaki tabloda iki değişken X ve Y için n=8 gözlem sayılı örneklem verileri için Spearman'ın sıralama korelasyon katsayısı ρ hesaplanması için örneğin verilmektedir. [A] ve [B] sütunlarında bu iki değişken X ve Y için örneklem verileri verilmiştir. [C] ve [D] sütunlarında bu iki değişkenlerin verileri için ayrı ayrı sıralama düzeni uygulanıp sıra numaraları x ve y olarak verilmiştir. X için verilerde 2 değişik beraberlik görülmektedir: 3 ve 10. Bu nedenle iki tekrarlı 3 için verilen sıra numaraları ortalaması (2+3)/2= 2,5 dur. Aynı şeklide 2 tekrarlı 10 için sıra numaraları 7,5 7,5 olarak verilmiştir. Y için verilerde ise 1,5 için 2 beraberlik ve 5 için 2 beraberlik bulunmaktadır ve bunlara da ortalama sıra numaraları verilmiştir. Sütun [E]de sıra numaraları farkları d verilmekte ve son [F] sütununda fark kareleri d^2 hesaplanmaktadır.

[A]	[B]	[C]	[D]	[E]	[F]
X	Y	x : X için sıralama	y : Y için sıralama	d : Sıralama farkları	d^2 : Farkların karesi
2	1,5	1	2,5	-1,5	2,25
3	1,5	2,5	2,5	0	0
3	4	2,5	5	-2,5	6,25
5	3	4	4	0	0
5,5	1	5	1	4	16
8	5	6	6,5	-0,5	0,25
10	5	7,5	6,5	1	1
10	9,5	7,5	8	-0,5	0,25
				Kareler Toplamı	26

$$\rho = 1 - \frac{6x26}{8(8^2 - 1)}$$

ρ=0.6 bulunur.

Bu ρ=0.6 değeri sıfıra yakın pozitiftir. Sıfıra yakınlığı X ve Y sıralamaları arasındaki bağlantının (korelasyonun) az olduğunu gösterir ve negatif olma ise var zayıf bağlantının aksi yönde olduğunu ifade eder (yani X sıralaması artarsa Y sıralaması düşer ve aksi olur).

Bu veriler içinde beraberlikler bulunmaktadır. Bu nedenle kullanılan genel ρ formülü uygun sonuç vermeyebilir. Daha uygun sonuç bulmak için x ve y sıra numaraları için Pearson'un çarpım-moment korelasyon katsayısı bulunması tavsiye edilmektedir.

3. Interpolasyon

İnterpolasyon, uygulamalı matematiğin bir dalı olan sayısal analiz yöntemlerini kullanılarak farklı bir yerde ve değeri bilinmeyen bir noktadaki olası değeri bulmaya ya da tahmin etmeye yarayan yöntemlerin tümüne verilen genel isimdir. En basit tanımı ile "varolan sayısal değerleri kullanarak, boş noktalardaki değerlerin tahmin edilmesi" olarak açıklanmaktadır. Türkçede bazen kolaylık olsun diye "interpolasyon" sözcüğü yerine yalnızca "tahmin" de kullanılmaktadır.

İnterpolasyon kavramı, verilen bir fonksiyon sınıfından, grafiği verilen sınırlı sayıdaki veri noktasından geçecek şekilde bir y=p(x) fonksiyonu seçme işlemidir. Kestirilen çıkış değeri çok boyutlu enterpolasyon ile sağlanmaktadır. Sistemin öğrenmesi ise, kestirilmiş değer ile yapılan kullanım sonrası ortaya çıkan hataya dayanarak gerçekleştirilir. Makine öğrenmesinde, hataların minimize edilmesi, temel kuraldır. **Önceden kestirilmiş olan yaklaşık değerler ile karşılaştırma yapılarak düzeltmeler gerçekleştirilir ve öğrenme sağlanır**. Çeşitli elemanların karakteristiklerinin otomatik olarak çıkartılması ve zamanla değişimlerinin izlenmesi yapılmalıdır. İzlenen sistem zamanla değişen bir sistem ise, algoritma öğrenmeye devam ederek, oluşturduğu fonksiyondaki katsayılarda gerekli değişiklikleri yapacaktır.

Bir fonksiyonun *xi* (*i=0,1,...N*) noktalarında bilinen *fi* (*i=0,1,...N*) değerlerinden hareketle herhangi bir *xo* ara noktasında bilinmeyen *f* (*xo*) ara değerinin bulunması anlamına gelen interpolasyon teknikleri aynı zamanda sayısal türev, integrasyon, adi ve kısmi türevli diferansiyel denklemlerin sayısal çözümü gibi başka sayısal yöntemlerin de esasını teşkil eder.

İnterpolasyon yöntemleri genellikle mevcut (xi , fi) veri noktalarına eğri veya eğriler uydurulması yoluyla uygulanır. Bu amaçla kullanılan fonksiyonlara interpolasyon fonksiyonları denir.

İnterpolasyon fonksiyonu olarak çoğu zaman çeşitli mertebeden (genellikle:,1,2,3) polinomlar kullanılır. Ancak bazı hallerde logaritmik, eksponansiyel, hiperbolik gibi daha özel fonksiyonlar, periyodik veri değerleri için trigonometrik fonksiyonlar kullanılabilir.

Veri noktaları eşit aralıklı olarak dağılmışsa sonlu fark esaslı interpolasyon yöntemleri, eşit aralıklı değilse doğrusal interpolasyon, Lagrange interpolasyonu vb yöntemler daha uygun olur.

İnterpolasyon polinomları:

Xi	3.2	2.7	1	4.8
yi	22	17.8	14.2	38.3

Bu noktalardan $y(x) = a_0 + a_1 x + a_2 x^2 + a3\, x^3$şeklinde üçüncü dereceden bir polinom (kübik) geçirmek mümkündür. Herbir noktanın koordinatları bu denklemi sağlayacağı için

$$a_0 + (3.2)a_1 + (3.2)^2 a_2 + (3.2)^3 a_3 = 22.0$$
$$a_0 + (2.7)a_1 + (2.7)^2 a_2 + (2.7)^3 a_3 = 17.8$$
$$a_0 + (1.0)a_1 + (1.0)^2 a_2 + (1.0)^3 a_3 = 14.2$$
$$a_0 + (4.8)a_1 + (4.8)^2 a_2 + (4.8)^3 a_3 = 38.3$$

$$\begin{bmatrix} 1 & 3.2 & 10.24 & 32.768 \\ 1 & 2.7 & 7.29 & 19.683 \\ 1 & 1.0 & 1.00 & 1.000 \\ 1 & 4.8 & 23.04 & 110.592 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{Bmatrix} = \begin{Bmatrix} 22.0 \\ 17.8 \\ 14.2 \\ 38.3 \end{Bmatrix}$$

Ax=b, linner denklem sisteminde, $x=A^{-1}b$
$a_0=24.3499;\ a_1=16.1177\ a_2=6.4952\ a_3=0.5275$

Yani interpolasyon fonksiyonu,
$y(x) = 24.3496 - 16.1176x + 6.4952x^2 - 0.5275x^3$

Buna göre örneğin *x=3.0* noktasındaki ara değer için *y=20.212* elde edilir.

Makine öğrenmesinde, verilen veri setinden öğrenen bir yapı elde edildi. Ardından test edilerek performansı artırılır. Performans artımı devamlılığı süreklidir.

4. Regresyon

En küçük kareler yöntemi, birbirine bağlı olarak değişen iki fiziksel büyüklük arasındaki matematiksel bağlantıyı, mümkün olduğunca gerçeğe uygun bir denklem olarak yazmak için kullanılan, standart bir regresyon yöntemidir. Bir başka deyişle bu yöntem, ölçüm sonucu elde edilmiş veri noktalarına "mümkün olduğu kadar yakın" geçecek bir fonksiyon eğrisi bulmaya yarar. Gauss-Markov Teoremi'ne göre en küçük kareler yöntemi, regresyon için optimal yöntemdir.

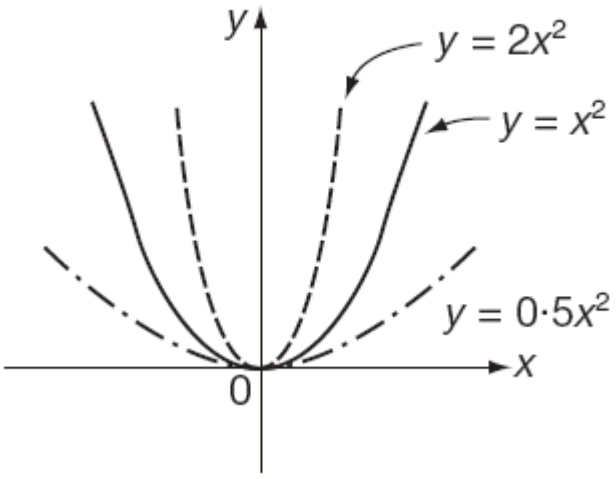
The variance of $\{x_1, \ldots, x_N\}$, denoted by σ_x^2, is

$$\sigma_x^2 = \frac{1}{N}\sum_{n=1}^{N}(x_i - \overline{x})^2;$$

the standard deviation σ_x is the square root of the variance:

$$\sigma_x = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(x_i - \overline{x})^2}.$$

Standard curves - *Second-degree curves*



The simplest second-degree curve is expressed by:

$$y = x^2$$

Its graph is a parabola, symmetrical about
The y-axis and existing only for $y \geq 0$. $y = ax^2$ gives a thinner parabola if $a > 1$ and a flatter parabola if $0 < a < 1$. The general second-degree curve is:

$$y = ax^2 + bx + c$$

where *a*, *b* and *c* determine the position,'width' and orientation of the parabola.

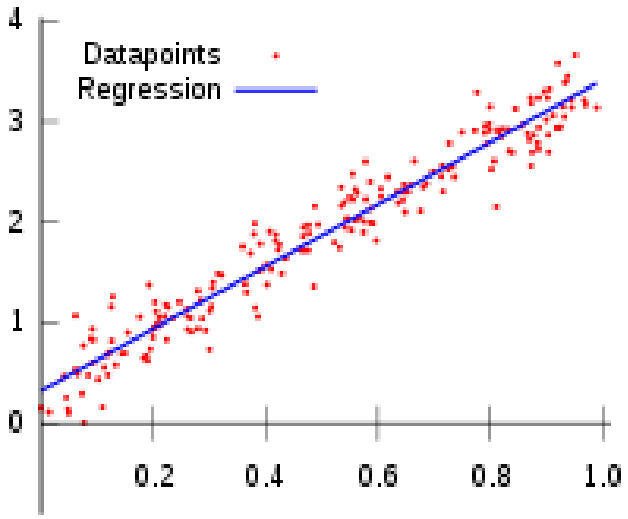
Linear Regression: $y(x) = ax + b$

Polynomial Regression: $y(x) = a_0 x^n + a_1 x^{n-1} + \cdots + a_{n-1} x + a_n$

1.order (linear): $y(x) = ax + b$

2.order: $y(x) = ax^2 + bx + c$

4.1. Doğrusal (Lineer) Regresyon



Kırmızı noktalar ölçümle elde edilmiş veri noktalarını, mavi çizgi ise en küçük kareler yöntemi ile bulunmuş teorik bağlantıyı ifade eder.

Çoğu zaman veri tablosuna tam olarak uyan bir fonksiyon bulmak mümkün olmaz; veri tablosuna en iyi uyan fonksiyon belirlenmeye çalışılır. Bir veri tablosuna en iyi uyan fonksiyonu bulma sürecine **regresyon analizi** denir. Regresyon analizi yaparken en çok kullanılan yöntemlerden biri en küçük kareler yöntemidir.

Belli ölçümler sonucunda i = 1, 2, . . . , n için (x_i , y_i) verileri elde edilmiş olsun. Burada, her bir y_i'nin x_i'ye bağlı olarak değiştiği varsayılmaktadır. Yapılan ölçümlerin doğası gereği, her i = 1, 2, . . . , n için $y_i = f(x_i)$ olacak biçimde bir fonksiyonun var olduğu, ölçümlerde yapılan hata nedeniyle bu eşitliklerin bazıları veya hepsinin sağlanmadığı kabul edilebilir. Bu düşünceyle, ölçülen y_i değeri $y(x_i)$ için yaklaşık değer kabul edilerek bu yaklaşımdaki hatanın minimum olduğu y fonksiyonu belirlenmeye çalışılır. Bu amacı gerçekleştirmek için f fonksiyonunun bir takım parametrelere bağlı bir ifadesi bulunduğu varsayılıp eldeki veriler yardımıyla bu parametreler belirlenmeye çalışılır. Örneğin, y fonksiyonu y = y (x) = mx + b ifadesinde olduğu gibi bir doğrusal fonksiyon veya $y = f(x) = ax^2 + bx + c$ ifadesinde olduğu gibi bir karesel fonksiyon olabilir ki bu durumda belirlenmesi gereken parametereler a , b , c , m dir.

En küçük kareler yönteminde aranan fonksiyon, ya da onun parametreleri, tüm farkların kareleri toplamı olan

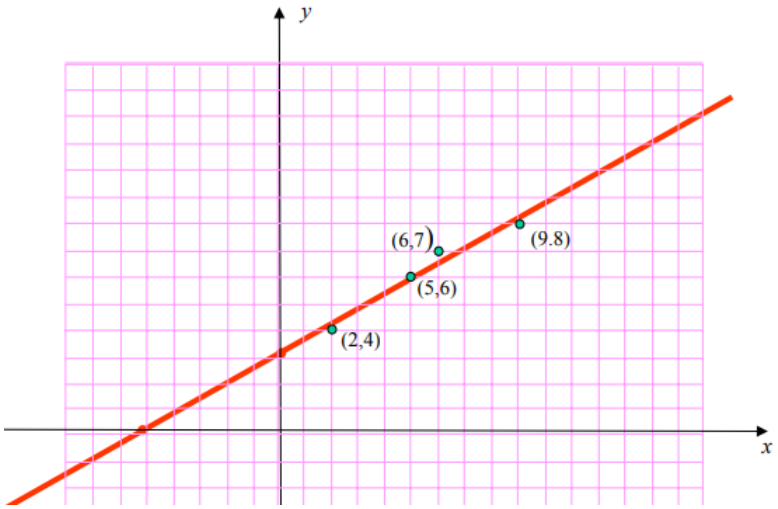
$$\sum_{i=1}^{n}\left(y_i-\hat{y}_i\right)^2=\sum_{i=1}^{n}\hat{\varepsilon}_i^2$$

ifadesini minimum yapacak şekilde belirlenir. Sözü edilen kareler toplamının minimum olması için her bir hatanın küçük olması gerekir. En Küçük Kareler, Kare Farklarının Toplamını (SSE) en aza indirir (Sum of the Squared Differences: SSE)

Örnek:

Bir makineden üretilen ürünlerin tablosu şu şekilde tutulmuştur,

2. ayda 4 adet
5. ayda 6 adet
6. ayda 7 adet
9.ayda 8 adet



f(x) = mx + b

$y_i = f(x_i)$

$$\sum_{i=1}^{n}(y_i - f(x_i))^2 = (y_1 - f(x_1))^2 + \cdots + (y_n - f(x_n))^2$$

2. ayda 4 adet , y1-f(x1)=4-2m-b
5. ayda 6 adet, y2-f(x2)=6-5m-b
6. ayda 7 adet, y3-f(x3)=7-6m+b
9.ayda 8 adet, y4-f(x4)=8-9m+b

$$F(m,b)=(4-2m-b)^2 + (6-5m-b)^2 + (7-6m-b)^2 + (8-9m-b)^2.$$

Eğimlerini bulabilmek için kısmi türevleri alınır ve sıfıra eşitlenir.

$$F_m(m,b)=2(4-2m-b)(-2)+2(6-5m-b)(-5)+2(7-6m-b)(-6)+2(8-9m-b)(-9)=0,$$

$$F_b(m,b)=2(4-2m-b)(-1)+2(6-5m-b)(-1)+2(7-6m-b)(-1)+2(8-9m-b)(-1)=0.$$

Sadeleştirmelerden sonra aşağıdaki denklemler elde edilir.
146m +22b=152
22m + 4b=25

Bu iki bilinmeyenli iki denklemin çözümünden m=0.58, b=3.06 bulunur.

f(x) = 0.58x + 3.06

m ve b nin doğrudan hesaplanmasını sağlayacak formüller farkların karelerinin toplamında elde edilir.

$$F(m,b)=\sum_{i=1}^{n}(y_i-mx_i-b)^2=(y_1-mx_1-b)^2+\cdots+(y_n-mx_n-b)^2$$

Yukarıdaki denklemde m ve b ye göre kısmi türevler alınıp sıfıra eşitlenirse aşağıdaki denklemler elde edilir.

$$F_m(m,b)=\sum_{i=1}^{n}2(y_i-mx_i-b)(-x_i)=-2\left(\sum_{i=1}^{n}x_i^2\right)m-2\left(\sum_{i=1}^{n}x_i\right)b+2\left(\sum_{i=1}^{n}x_iy_i\right)=0$$

$$F_b(m,b)=\sum_{i=1}^{n}2(y_i-mx_i-b)(-1)=-2\left(\sum_{i=1}^{n}x_i\right)m-2\left(\sum_{i=1}^{n}1\right)b+2\left(\sum_{i=1}^{n}y_i\right)=0$$

ya da

$$\begin{cases}\left(\sum_{i=1}^{n}x_i^2\right)m+\left(\sum_{i=1}^{n}x_i\right)b=\sum_{i=1}^{n}x_iy_i\\ \left(\sum_{i=1}^{n}x_i\right)m+\quad nb\quad=\sum_{i=1}^{n}y_i\end{cases}$$

denklem sistemi çözülerek bulunur.

Bu denklem sisteminin daima tek bir çözümü bulunduğuna dikkat ediniz.

$$m = \frac{n(\sum_{k=1}^{n} x_k y_k) - (\sum_{k=1}^{n} x_k)(\sum_{k=1}^{n} y_k)}{n(\sum_{k=1}^{n} x_k^2) - (\sum_{k=1}^{n} x_k)^2} \quad , \quad b = \frac{\sum_{k=1}^{n} y_k - m(\sum_{k=1}^{n} x_k)}{n}.$$

Örnek:

(0 , 6.4), (1 , 2.6), (2 , 0.5), (3 , 0.6) ve (4 , 0.3) veri noktalarına en iyi uyan y=mx + b, doğru denklemini bulunuz.

$$\sum_{k=1}^{5} x_k = 0+1+2+3+4 = 10 \quad , \quad \sum_{k=1}^{5} y_k = 6.4+2.6+0.5+0.6+0.3 = 10.4$$

$$\sum_{k=1}^{5} x_k y_k = 0\cdot(6.4)+1\cdot(2.6)+2\cdot(0.5)+3\cdot(0.6)+4\cdot(0.3) = 2.6+1+1.8+1.2 = 6.6$$

$$\sum_{k=1}^{5} x_k^{\,2} = 0+1+4+9+16 = 30 \quad , \quad (\sum_{k=1}^{5} x_k)(\sum_{k=1}^{5} y_k) = 10\cdot(10.4) = 104$$

$$m = \frac{n(\sum_{k=1}^{n} x_k y_k) - (\sum_{k=1}^{n} x_k)(\sum_{k=1}^{n} y_k)}{n(\sum_{k=1}^{n} x_k^2) - (\sum_{k=1}^{n} x_k)^2} = \frac{5\cdot(6.6)-104}{5\cdot 30-100} = \frac{33-104}{50} = \frac{-71}{50} = -1.42,$$

$$b = \frac{\sum_{k=1}^{n} y_k - m(\sum_{k=1}^{n} x_k)}{n} = \frac{10.4-(-1.42)\cdot 10}{5} = \frac{10.4+14.2}{5} = \frac{24.6}{5} = 4.92.$$

y=-1.42x+4.92

4.2. İkinci Mertebe En Küçük Kareler Yöntemi

$$E = \sum_{i=1}^{n} (y_i - f(x_i))^2$$

$f(x)=a_2x^2 + a_1 x + a_0$

$$E = \sum_{i=1}^{N} (y_i - a_2x_i^2 - a_1bx_i - a_0)^2$$

$E_a = \sum_{i=1}^{N} 2(y_i - a_2x_i^2 - a_1x_i - a_0)(-x_i^2)=0$

$E_b = \sum_{i=1}^{N} 2(y_i - a_2x_i^2 - a_1x_i - a_0)(-x_i)=0$

$E_c = \sum_{i=1}^{N} 2(y_i - a_2x_i^2 - a_1x_i - a_0)(-1)=0$

$$\sum_{i=1}^{N} y_i x_i^2 = a_2 \sum_{i=1}^{N} x_i^4 + a_1 \sum_{i=1}^{N} x_i^3 + a_0 \sum_{i=1}^{N} x_i^2$$

$$\sum_{i=1}^{N} y_i x_i = a_2 \sum_{i=1}^{N} x_i^3 + a_1 \sum_{i=1}^{N} x_i^2 + a_0 \sum_{i=1}^{N} x_i$$

$$\sum_{i=1}^{N} y_i = a_2 \sum_{i=1}^{N} x_i^2 + a_1 \sum_{i=1}^{N} x_i + a_0 N$$

4.3. Genel Polinom Regresyon Modeli

Genel polinom regresyon modeli, en küçük kareler yöntemi kullanılarak geliştirilebilir. En küçük kareler yöntemi, polinomdan tahmin edilen değerler ile veri kümesinden beklenen değerler arasındaki farkı en aza indirmeyi amaçlamaktadır.

The coefficients of the polynomial regression model (a_k,a_{k-1},⋯,a_1, a_0) may be determined by solving the following system of linear equations.

$$\begin{bmatrix} N & \sum_{i=1}^{N} x_i & \cdots & \sum_{i=1}^{N} x_i^k \\ \sum_{i=1}^{N} x_i & \sum_{i=1}^{N} x_i^2 & \cdots & \sum_{i=1}^{N} x_i^{k+1} \\ \vdots & \vdots & \vdots & \vdots \\ \sum_{i=1}^{N} x_i^k & \sum_{i=1}^{N} x_i^{k+1} & \cdots & \sum_{i=1}^{N} x_i^{2k} \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_k \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{N} y_i \\ \sum_{i=1}^{N} x_i y_i \\ \vdots \\ \sum_{i=1}^{N} x_i^k y_i \end{bmatrix}$$

Örnek:
Aşağıdaki veri kümesine uygun 2. dereceden bir polinom eğrisinin nasıl geliştirileceğini gösterin.

x	-3	-2	-1	-0.2	1	3
y	0.9	0.8	0.4	0.2	0.1	0

Bu veri kümesinde N = 6 ve 2. dereceden bir polinom için k = 2 dir. En küçük kareler yönteminin uygulanması aşağıdaki doğrusal sistemi sağlar.

$$\begin{bmatrix} 6 & -2.2 & 24.04 \\ -2.2 & 24.04 & -8.008 \\ 24.04 & -8.008 & 180.0016 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} 2.4 \\ -4.64 \\ 11.808 \end{bmatrix}$$

Sistemi çözmek için Cramer kuralını kullanarak, M matrisini alarak ve sütun vektörü b'yi i sütununa yerleştirerek Mi matrislerinin her birini üretiyoruz, örneğin M0, M1, M2,:

$$M_0 = \begin{bmatrix} 2.4 & -2.2 & 24.04 \\ -4.64 & 24.04 & -8.008 \\ 11.808 & -8.008 & 180.0016 \end{bmatrix}$$

$$M_1 = \begin{bmatrix} 6 & 2.44 & 24.04 \\ -2.2 & -4.64 & -8.008 \\ 24.04 & 11.808 & 180.0016 \end{bmatrix}$$

$$M_2 = \begin{bmatrix} 6 & 2.44 & 2.4 \\ -2.2 & -4.64 & -4.64 \\ 24.04 & 11.808 & 11.808 \end{bmatrix}$$

$$a_0 = \frac{det(M_0)}{det(M)} = \frac{2671.20}{11661.27} = 0.2291$$
$$a_1 = \frac{det(M_1)}{det(M)} = \frac{-1898.46}{11661.27} = -0.1628$$
$$a_2 = \frac{det(M_2)}{det(M)} = \frac{323.76}{11661.27} = 0.0278$$

$$y = 0.0278x^2 - 0.1628x + 0.2291$$

Örnek:

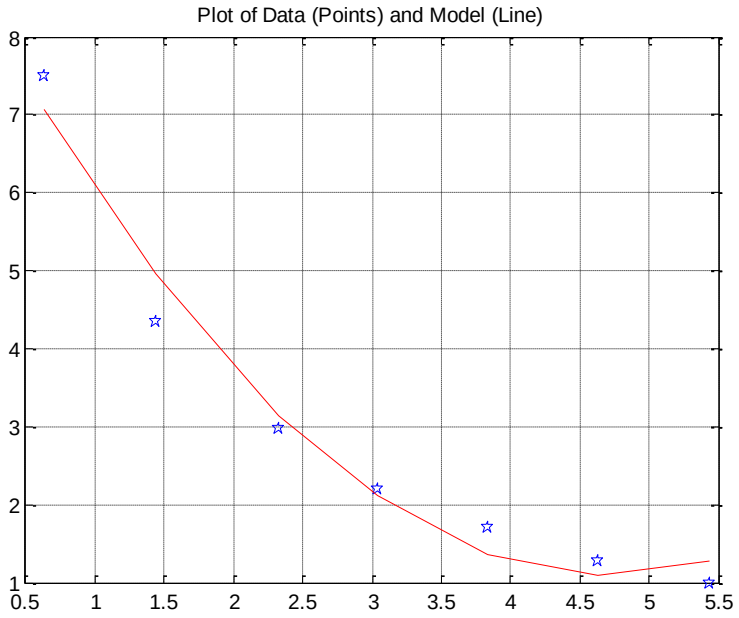
ikinci dereceden en küçük kareler ve Matlab ; p = polyfit(x,y,n), n=2

```
clear all
close all
x = [5.435   4.635   3.835   3.035   2.325   1.435   0.635];
y = [1.00     1.28     1.70     2.20     2.97     4.35     7.50  ];
p = polyfit(x, y, 2)                        % Quadratic Function Fit
v = polyval(p, x)                           % Evaluate
TSE = sum((v - y).^2)                % Total Squared Error
figure(1)
plot(x, y, 'bp')
hold on
plot(x, v, '-r')
hold off
grid
title('Plot of Data (Points) and Model (Line)')
```

p = 0.3582 -3.3833 9.0748

v = 1.2664 1.0877 1.3674 2.1056 3.1447 4.9573 7.0708

TSE = 0.8110



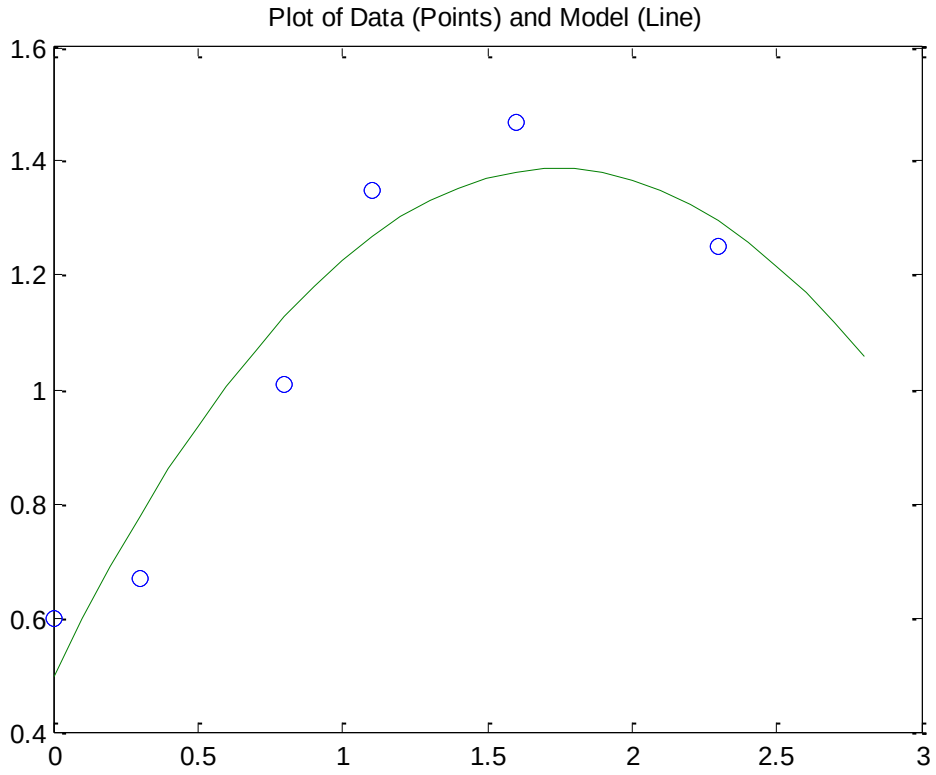
Örnek:

ikinci dereceden en küçük kareler ve Matlab ; p = polyfit(x,y,n), n=2

```
clear all
close all
t = [0 0.3 0.8 1.1 1.6 2.3];
y = [0.6 0.67 1.01 1.35 1.47 1.25];
figure, plot(t,y,'o')
title('Plot of y Versus t')
p = polyfit(t,y,2)
t2 = 0:0.1:2.8;
y2 = polyval(p,t2);
hold on
plot(t,y,'o',t2,y2)
title('Plot of Data (Points) and Model (Line)')
```

p = -0.2942 1.0231 0.4981

f(x) = $-0.2942x^2 + 1.0231x + 0.4981$



Örnek:

Uyum İşlevini Kullanarak Üstel Modellere Sığdırma

```
clear all
close all
x = (0:0.2:5)';
y = 2*exp(-0.2*x) + 0.1*randn(size(x));
f = fit(x,y,'exp1')
plot(f,x,y)
```

f = General model Exp1:

f(x) = a*exp(b*x)

Coefficients (with 95% confidence bounds):

a = 2.103 (1.999, 2.208)

b = -0.2222 (-0.2464, -0.1981)