Makine Öğrenmesi Ders Notları
Hazırlayan: Barış Baykant Alagöz
İnönü Üniversitesi, Bilgisayar Mühendisliği, 2023-2024 Bahar Dönemi Lisans Dersi
Sorularınız ve düzeltme önerileriniz için e-mail: baykant.alagoz@inonu.edu.tr

Regresyon Problemleri:

En Küçük Kareler Yöntemi ile Kuadratik(Quadratic) Regresyon

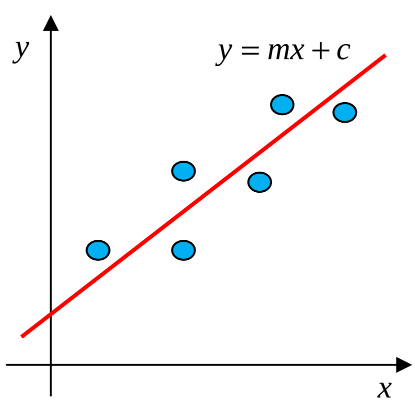
Uygulamada her zaman lineer bir modeller yeterli olmayabilir veya verileri yeterince iyi modelleyemeyebilir. Veriler lineer(doğrusal) bir karakterde olmayabilir ve model karmaşıklığının artırılmasına ihtiyaç duyulabilir. Polinomlar karmaşıklığı daha yüksek ve lineer olmayan karakterdeki verileri modelleyebilir. Bu derste polinomsal modellere örnek olarak 2. derece polinom olan kuadratik fonksiyonlar (paraboller) ile modelleme incelenecektir.

Lineer model: $y = m_1 x + c$

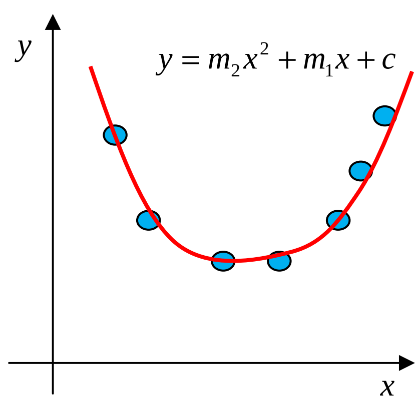
Kuadratik model: $y = m_2 x^2 + m_1 x + c$

Burada lineer modeli kuadratik model yapmak için eklenen $m_2 x^2$ terimi hem model karmaşıklığını artırır hemde bu lineer olmayan terim lineer olmayan kuadratik (2. kuvvete bağlı) ilişkileri modelleyebilir.

[Polinomsal modeller n. derece polinomsal fonksiyonlarına genişletilebilir. $y = m_n x^n + m_{n-1} x^{n-1} + ... + m_2 x^2 + m_1 x + c$. Ancak dersimizin içeriğinde bu tür modeller için işlem yükü fazlalılığından dolayı manüel çözümleme yapılmayacak. Kuadratik modeller ile çalışılacaktır. Ancak, uygulana çözüm, n. derece polinomsal fonksiyonlarına genişletilebilir.]



Doğrusal karakterde veriler, bir lineer model yardımı ile modellenebilir.



Parabolik karakterde veriler, bir kuadratik model yardımı ile daha iyi modellenebilir.

Varsayalım, $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), ..., (x_n, y_n)$ verileri $y = m_2 x^2 + m_1 x + c$ ile modellenmek istensin. olsun. Bu problemde bu modeli elde etmek için verilen $(x_1, y_1), (x_2, y_2), (x_3, y_3), (x_4, y_4), ..., (x_n, y_n)$ n adet noktayı sağlayan $y = m_2 x^2 + m_1 x + c$

modelinin m_1, m_2 ve c parametrelerini bulmalıyız. Bu noktaları denklemde kullanarak denklem sistemini elde edelim:

(x_1, y_1) verisi için model sağlatılırsa $y_1 = m_2 x_1^2 + m_1 x_1 + c$ denklemi

(x_1, y_1) verisi için model sağlatılırsa $y_2 = m_2 x_2^2 + m_1 x_2 + c$

....

(x_n, y_n) sağlatılırsa $y_n = m_2 x_n^2 + m_1 x_n + c$ denklemi elde edilir. Böylece bütün veri noktalarının sağlanması için

$$y_1 = m_2 x_1^2 + m_1 x_1 + c$$
$$y_2 = m_2 x_2^2 + m_1 x_2 + c$$
.....
$$y_n = m_2 x_n^2 + m_1 x_n + c$$

n adet 3 bilinmeyenli denklem sistemi çözülmesi gerekmektedir. Bilinmeyen sayısı 3 adettir (m_2, m_1, c) ancak denklem sayısı $n > 3$ olduğunu varsayınız. Karşımıza aslında aşırı belirlenmiş bir denklem sistemi çıkar ve lineer regresyon olduğu gibi en küçük kareler yöntemi ile çözülür.

Denklem sistemi matris formunda yazılırsa,

$$\begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ \dots & \dots & \dots \\ x_n^2 & x_n & 1 \end{bmatrix} \begin{bmatrix} m_2 \\ m_1 \\ c \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ .. \\ y_n \end{bmatrix} \Rightarrow A = \begin{bmatrix} x_1^2 & x_1 & 1 \\ x_2^2 & x_2 & 1 \\ \dots & \dots & \dots \\ x_n^2 & x_n & 1 \end{bmatrix}, X = \begin{bmatrix} m_2 \\ m_1 \\ c \end{bmatrix}, b = \begin{bmatrix} y_1 \\ y_2 \\ .. \\ y_n \end{bmatrix}$$

denklem sistemi matris formunda $A.X = b$ elde edilir.

Burada A matrisi kare matris olamaz ve tersi alınamaz. Tersinin alınabilmesi için kareselleştirme yapılmalı. $(A^T.A)$ kare matristir. Eşitliğin her iki tarafı A^T ile çarpılırsa,

$(A^T.A).X = A^T.b$ sonra $(A^T.A)^{-1}$ ile çarpılırsa,

$$(A^T.A)^{-1}(A^T.A).X = (A^T.A)^{-1} A^T.b$$

çözüm $X = (A^T.A)^{-1} A^T.b$ elde edilir.

Aşırı belirlenmiş sistem formundaki lineer problemlerin en küçük kareler çözümü,

$$\boxed{X = (A^T.A)^{-1} A^T.b}$$

ile elde edilir.

Örnek: Kayısı üretiminin yağış oranına göre dağılımı şöyle olsun.

Kayısı Üretimi (K)	Yağış Oranı (Y)
9	1
6	2
7	3
10	4

verilerini kuadratik regresyon modelini $K = m_2 Y^2 + m_1 Y + c$ elde ediniz. Tablodaki veriler $K = m_2 Y^2 + m_1 Y + c$ 'de kullanılırsa,

$9 = m_2 1^2 + m_1 1 + c$, => $9 = m_2 + m_1 + c$

$6 = m_2 2^2 + m_1 2 + c$ => $6 = m_2 4 + m_1 2 + c$

$7 = m_2 3^2 + m_1 3 + c$ => $7 = m_2 9 + m_1 3 + c$

$10 = m_2 4^2 + m_1 4 + c$ => $10 = m_2 16 + m_1 4 + c$

lineer denklemleri elde edilir. Bu lineer denklemleri sağlayan minimum kareler çözümü,

$X = (A^T.A)^{-1} A^T.b$ ile bulunur. Bunun için önce denklemleri matris formunda yazalım.

$$\begin{bmatrix} 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \\ 16 & 4 & 1 \end{bmatrix} \begin{bmatrix} m_2 \\ m_1 \\ c \end{bmatrix} = \begin{bmatrix} 9 \\ 6 \\ 7 \\ 10 \end{bmatrix}$$

=> $A = \begin{bmatrix} 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \\ 16 & 4 & 1 \end{bmatrix}, X = \begin{bmatrix} m_2 \\ m_1 \\ c \end{bmatrix}, b = \begin{bmatrix} 9 \\ 6 \\ 7 \\ 10 \end{bmatrix}$

$(A^T.A) = \begin{bmatrix} 354 & 100 & 30 \\ 100 & 30 & 10 \\ 30 & 10 & 4 \end{bmatrix}$ ve $A^T.b = \begin{bmatrix} 256 \\ 82 \\ 32 \end{bmatrix}$ elde edilir. Kare matris tersi alınabilir

$$(A^T.A)^{-1} = \begin{bmatrix} 0.25 & -1.25 & 1.25 \\ -1.25 & 6.54 & -6.75 \\ 1.25 & -6.75 & 7.75 \end{bmatrix}$$

$$X = \begin{bmatrix} m_2 \\ m_1 \\ c \end{bmatrix} \Rightarrow X = (A^T.A)^{-1}A^T.b = \begin{bmatrix} 354 & 100 & 30 \\ 100 & 30 & 10 \\ 30 & 10 & 4 \end{bmatrix} . \begin{bmatrix} 256 \\ 82 \\ 32 \end{bmatrix} = \begin{bmatrix} 1.5 \\ -7.1 \\ 14.5 \end{bmatrix}$$

dolayısı ile $m_2 = 1.5$, $m_1 = -7.1$ ve $c = 14.5$ ve kuadratik regresyon modeli $K = 1.5Y^2 - 7.1Y + 14.5$

Yağış tahmini uygulaması:

Bu model göre $Y = 3.2$ yağış miktarı için kayısı üretimini tahmin ediniz. Giriş değişkenimiz yağış miktarı ve çıkış değişkeni kayısı miktarı tahminidir.

$K = 1.5(3.2)^2 - 7.1(3.2) + 14.5 = 7.14$ olarak tahmin edilir.

Şimdi bu regresyon problemi çözümü için karesel hatayı hesaplayalım.

Yağış Oranı (Y)	Kayısı Üretimi (K) (istenen değer)	Öğrenen model tahmini $K = 2Y + 3$	Öğrenme Hatası $e = K - (2Y + 3)$
1	9	8.9	0.1
2	6	6.3	0.3
3	7	6.7	-0.3
4	10	10.1	0.1

$E = \frac{1}{4}\sum_{i=1}^{4} e_i^2 = \frac{1}{4}(0.01 + 0.09 + 0.09 + 0.01) = 0.05$ ortalama karesel hata sıfırdır.

Örnek1: Önceki örnekte verilen $K = 2Y + 3$ ile hatasız modellenebilen lineer karakterdeki kayısı üretiminin yağış oranına göre dağılımı verileri $K = m_2Y^2 + m_1Y + c$ ile modellenseydi ne beklenirdi?

Kayısı Üretimi (K)	Yağış Oranı (Y)
5	1
7	2
9	3
11	4

Tablodaki veriler $K = m_2Y^2 + m_1Y + c$ 'de kullanılırsa,

$5 = m_2 1^2 + m_1 1 + c$, $\Rightarrow$ $5 = m_2 + m_1 + c$

$7 = m_2 2^2 + m_1 2 + c \Rightarrow 7 = m_2 4 + m_1 2 + c$

$9 = m_2 3^2 + m_1 3 + c \Rightarrow 9 = m_2 9 + m_1 3 + c$

$11 = m_2 4^2 + m_1 4 + c$ => $11 = m_2 16 + m_1 4 + c$

lineer denklemleri elde edilir. Bu lineer denklemleri sağlayan minimum kareler çözümü,

$X = (A^T.A)^{-1} A^T.b$ ile bulunur. Bunun için önce denklemleri matris formunda yazalım.

$$\begin{bmatrix} 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \\ 16 & 4 & 1 \end{bmatrix} \begin{bmatrix} m_2 \\ m_1 \\ c \end{bmatrix} = \begin{bmatrix} 5 \\ 7 \\ 9 \\ 11 \end{bmatrix}$$

$$\Rightarrow A = \begin{bmatrix} 1 & 1 & 1 \\ 4 & 2 & 1 \\ 9 & 3 & 1 \\ 16 & 4 & 1 \end{bmatrix}, X = \begin{bmatrix} m_2 \\ m_1 \\ c \end{bmatrix}, b = \begin{bmatrix} 5 \\ 7 \\ 9 \\ 11 \end{bmatrix}$$

$(A^T.A) = \begin{bmatrix} 354 & 100 & 30 \\ 100 & 30 & 10 \\ 30 & 10 & 4 \end{bmatrix}$ ve $A^T.b = \begin{bmatrix} 290 \\ 90 \\ 32 \end{bmatrix}$ elde edilir. Kare matris tersi alınabilir

$$(A^T.A)^{-1} = \begin{bmatrix} 0.25 & -1.25 & 1.25 \\ -1.25 & 6.54 & -6.75 \\ 1.25 & -6.75 & 7.75 \end{bmatrix}$$

$$X = \begin{bmatrix} m_2 \\ m_1 \\ c \end{bmatrix} \Rightarrow X = (A^T.A)^{-1} A^T.b = \begin{bmatrix} 354 & 100 & 30 \\ 100 & 30 & 10 \\ 30 & 10 & 4 \end{bmatrix} . \begin{bmatrix} 290 \\ 90 \\ 32 \end{bmatrix} = \begin{bmatrix} 0 \\ 3 \\ 2 \end{bmatrix}$$

dolayısı ile $m_2 = 0$, $m_1 = 3$ ve $c = 2$ ve kuadratik regresyon modeli $K = 0.Y^2 + 3Y + 2$ elde edilir. Buda aslında lineer regresyon modeli olan $K = 3Y + 2$ 'dir. Yani model değişmedi. Veriler lineer regresyon doğrusu ile hatasız modellendiği için kuadratik modelde optimizasyon sonucunda $m_2 = 0$ elde edilebildi. Kuadratik model bir lineer model dönüşebilir. Burada aynı zamanda kuadratik modellerin aslında lineer modelleride kapsayabildiğini görebiliriz. [Ancak, bu etki sadece sıfır karesel hatada görülebilir. Lineer modelin hatası sıfır olmasaydı $m_2 = 0$ olmayabilirdi.]