برازش منحنی

CURVE FITTING

برازش منحنی (CURVE FITTING)

در بسیاری از موارد از نتایج تجربی و آزمایشگاهی، یکسری داده های گسسته در اختیار داریم.

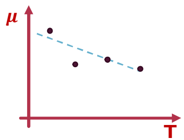
✓ استفاده از مدلی ریاضی بجای جدول برای محاسبات بعدی

مدل های ریاضی:

1. رگرسیون
2. درون یابی

رگرسیون

- ❑ عدم اطمینان به داده های تجربی و همراه بودن با خطاها
- ❑ عدم لزوم گذر مدل از تمامی نقاط



مدل ارائه شده (رگرسیون)، رفتار کلی نقاط را نشان میدهد.

❖ مثال: با افزایش دما، ویسکوزیته با شیبی کاهش یافته است.

Temp.	T_1	T_2	T_3	T_4
Viscosity	μ_1	μ_2	μ_3	μ_4

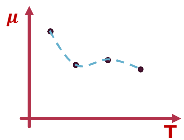
درونیابی

اگر به داده های تجربی مطمئن باشیم:

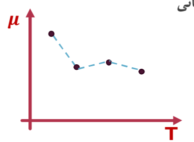
✓ گذر مدل ارائه شده از تمام نقاط

✓ امکان محاسبه مقدار تابع در نقاط میانی

- درونیابی خطی
- درونیابی غیرخطی

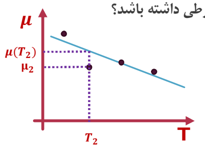


درونیابی غیرخطی



درونیابی خطی

رگرسیون



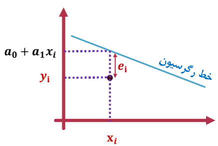
- اگر بخواهیم رفتار نقاط را با یک مدل خطی توصیف کنیم این خط چه شرطی داشته باشد؟
- آیا این خط، یکتا است یا به شرط ما بستگی دارد؟
- میزان این اختلاف، مقدار خطای مدل در نقطه است.
- یکی از معیارها: حداقل بودن این خطاها
- معیار حداقل سازی مجموع مربعات خطا (برای اینکه علامت خط مهم نباشد)
- به این معیار، رگرسیون حداقل مربعات **least square regression** گفته میشود.

انواع رگرسیون

1. رگرسیون خطی تک متغیره (**Linear least square regression**)
2. رگرسیون خطی چند متغیره (**multiple Linear least square regression**)
3. رگرسیون چندجمله ای تک متغیره (**polynomial least square regression**)

رگرسیون خطی تک متغیره (LINEAR LEAST SQUARE REGRESSION)

$y = a_0 + a_1 x$ مدل خطی رگرسیون



برای مدل خطی حداقل به دو نقطه نیاز داریم که معمولا بیشتر هستند. معمولا **n** نقطه داریم.

x	x_1	x_2	x_3	…	x_n
y	y_1	y_2	y_3	…	y_n

اختلاف بین مقدار تجربی و مقدار تجربی بدست آمده از معادله خط

$$e_i = y_i - (a_0 + a_1 x_i)$$

مجموع مربعات خطا

$$S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

حداقل سازی مجموع مربعات خطا

$$e_i = y_i - (a_0 + a_1 x_i) \qquad \Rightarrow S_r = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2$$

میخواهیم S_r را حداقل نماییم.

مجهولات چیست؟ a_1 و a_0

$$\Rightarrow \begin{cases} \dfrac{\partial S_r}{\partial a_0} = 0 \\ \dfrac{\partial S_r}{\partial a_1} = 0 \end{cases} \Rightarrow \begin{cases} -2\sum_{i=1}^{n}(y_i - a_0 - a_1 x_i) = 0 \\ -2\sum_{i=1}^{n} x_i (y_i - a_0 - a_1 x_i) = 0 \end{cases}$$

$$\Rightarrow \begin{cases} \sum y_i - \sum a_0 - \sum a_1 x_i = 0 \\ \sum x_i y_i - \sum a_0 x_i - \sum a_1 x_i^2 = 0 \end{cases} \Rightarrow \begin{cases} n a_0 + a_1 (\sum x_i) = \sum y_i \\ a_0 (\sum x_i) + a_1 (\sum x_i^2) = \sum x_i y_i \end{cases}$$

$$\Rightarrow \begin{cases} a_1 = \dfrac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \\ a_0 = \overline{y} - a_1 \overline{x} \end{cases} \qquad \overline{x} = \frac{\sum x_i}{n}, \overline{y} = \frac{\sum y_i}{n}$$

فرم ماتریسی دستگاه معادلات

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum {x_i}^2 \end{bmatrix}\begin{Bmatrix} a_0 \\ a_1 \end{Bmatrix}=\begin{Bmatrix} \sum y_i \\ \sum x_i y_i \end{Bmatrix}$$

میتوان از روشهای ماتریسی آن را حل کرد و مقادیر مجهول را بدست آورد.

$S_t=\sum_{i=1}^{n}(y_i-\bar{y})^2$

$S_y=\sqrt{\frac{S_t}{n-1}}$ انحراف استاندارد

s_y^2 واریانس

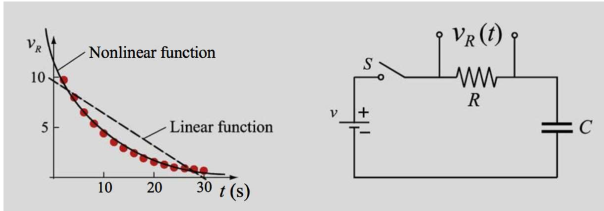
$r=\sqrt{\frac{S_t-S_r}{S_t}}$ ضریب همبستگی

$r^2=\frac{S_t-S_r}{S_t}$ ضریب تعیین

$$r^2=\mathrm{r}=1 \quad \rightarrow \quad S_r=0$$

هرچقدر به یک نزدیکتر باشند، یعنی خط رگرسیون به نقاط جدول نزدیک تر است.

تبدیل روابط به رگرسیون خطی



تبدیل روابط به رگرسیون خطی

- خیلی وقت ها میتوان مدل مدنظر را با تغییراتی به فرم مدل خطی نوشت:

$y=a_1e^{b_1x}$ Y بر حسب x خطی نیست.

$\ln y=\ln a_1+b_1\mathrm{x} \quad \Rightarrow \; y=b_0+b_1\mathrm{x}$ Ln(y) بر حسب x خطی است.

$y=a_1\frac{x}{b_1+x}$ Y بر حسب x خطی نیست.

$\frac{1}{y}=\frac{b_1}{a_1}\frac{1}{x}+\frac{1}{a_1} \qquad \mathrm{Y}=\mathrm{b_0X}+\mathrm{b_1}$ 1/y بر حسب 1/x خطی است.

$y=a_1x^{b_1}$ Y بر حسب x خطی نیست.

$\log y=\log a_1+b_1\log x \quad \Rightarrow Y=b_0+b_1\mathrm{X}$ Log(y) بر حسب log(x) خطی است.

رگرسیون خطی چند متغیره (MULTIPLE LINEAR LEAST SQUARE REGRESSION)

y_i	x_{1i}	x_{2i}	x_{3i}	…	x_{mi}
y_1	x_{11}	x_{21}	x_{31}	…	x_{m1}
y_2	x_{12}	x_{22}	x_{32}	…	x_{m2}
⋮	⋮	⋮	⋮	⋱	⋮
y_n	x_{1n}	x_{2n}	x_{3n}	…	x_{mn}

اگر n داده داشته باشیم (y تا n) که هرکدام تابع m تا x است.

$n\geq(m+1)$

$y^*=a_0+a_1x_1+a_2x_2+\cdots+a_mx_m$ تابع رگرسیون

مجهولات a_0 تا a_m

$e_i=y_i-y_i^*=y_i-(a_0+a_1x_{1i}+a_2x_{2i}+\cdots+a_mx_{mi})$ تابع خطا

$S_r=\sum_{i=0}^{n}e_i^2=\sum_{i=0}^{n}\left(y_i-(a_0+a_1x_{1i}+a_2x_{2i}+\cdots+a_mx_{mi})\right)^2$ مجموع مربعات خطا

رگرسیون خطی چند متغیره (MULTIPLE LINEAR LEAST SQUARE REGRESSION)

حال باید s_r نسبت به تمام متغیرها (a_0 تا a_m) حداقل شود.

$$\begin{cases} \dfrac{\partial s_r}{\partial a_0} = -2\sum\left(y_i - (a_0 + a_1x_{1i} + a_2x_{2i} + \cdots + a_mx_{mi})\right) = 0 \\ \dfrac{\partial s_r}{\partial a_1} = -2\sum(x_{1i})\left(y_i - (a_0 + a_1x_{1i} + a_2x_{2i} + \cdots + a_mx_{mi})\right) = 0 \\ \vdots \\ \dfrac{\partial s_r}{\partial a_m} = -2\sum(x_{mi})\left(y_i - (a_0 + a_1x_{1i} + a_2x_{2i} + \cdots + a_mx_{mi})\right) = 0 \end{cases}$$

$$\Rightarrow \begin{cases} \sum y_i - na_0 - a_1\sum x_{1i} - \cdots - a_m\sum x_{mi} = 0 \\ \sum y_i\,x_{1i} - a_0\sum x_{1i} - a_1\sum {x_{1i}}^2 - \cdots - a_m\sum x_{1i}x_{mi} = 0 \\ \vdots \\ \sum y_i\,x_{mi} - a_0\sum x_{mi} - a_1\sum x_{mi}x_{1i} - \cdots - a_m\sum {x_{mi}}^2 = 0 \end{cases}$$

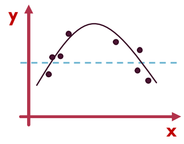
بنابراین به یک دستگاه معادلات خطی با m+1 معادله و m+1 مجهول رسیدیم.

فرم ماتریسی معادلات

$$\Rightarrow \begin{bmatrix} n & \sum x_{1i} & \sum x_{2i} & \cdots & \sum x_{mi} \\ \sum x_{1i} & \sum {x_{1i}}^2 & \sum x_{1i}x_{2i} & \cdots & \sum x_{1i}x_{mi} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sum x_{mi} & \sum x_{mi}x_{1i} & \sum x_{mi}x_{2i} & \cdots & \sum {x_{mi}}^2 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{Bmatrix} = \begin{Bmatrix} \sum y_i \\ \sum x_{1i}y_i \\ \vdots \\ \sum x_{mi}y_i \end{Bmatrix}$$

رگرسیون تک متغیره چندجمله ای (POLYNOMIAL LEAST SQUARE REGRESSION)

- در بسیاری از موارد، داده های مهندسی رفتاری خطی از خود نشان نمیدهند.
  - ✓ استفاده از توابع چندجمله ای
- تابع رگرسیون نسبت به x چندجمله ای است.
- نسبت به ضرایب a همچنان خطی رفتار میکند.



$$y^* = a_0 + a_1x + a_2x^2 + \cdots + a_mx^m$$

$n \ge (m+1)$

رگرسیون تک متغیره چندجمله ای

$$s_r = \sum_{i=0}^{n}\left(y_i - (a_0 + a_1x_i + a_2{x_i}^2 + \cdots + a_m{x_i}^m)\right)^2$$

مجموع مربعات خطا

حالا باید s_r نسبت به تمام متغیرها (a_0 تا a_m) حداقل شود.

$$\begin{cases} \dfrac{\partial s_r}{\partial a_0} = -2\sum\left(y_i - (a_0 + a_1x_i + a_2{x_i}^2 + \cdots + a_m{x_i}^m)\right) = 0 \\ \dfrac{\partial s_r}{\partial a_1} = -2\sum(x_i)\left(y_i - (a_0 + a_1x_i + a_2{x_i}^2 + \cdots + a_m{x_i}^m)\right) = 0 \\ \vdots \\ \dfrac{\partial s_r}{\partial a_m} = -2\sum({x_i}^m)\left(y_i - (a_0 + a_1x_i + a_2{x_i}^2 + \cdots + a_m{x_i}^m)\right) = 0 \end{cases}$$

رگرسیون تک متغیره چندجمله ای

$$\Rightarrow \begin{cases} \sum y_i - na_0 - a_1 \sum x_i - a_2 \sum x_i^2 - \cdots - a_m \sum x_i^m = 0 \\ \sum y_i x_i - a_0 \sum x_i - a_1 \sum x_i^2 - a_2 \sum x_i^3 - \cdots - a_m \sum x_i^m = 0 \\ \vdots \\ \sum x_i^m y_i - a_0 \sum x_i^m - a_1 \sum x_i^{m+1} - a_2 \sum x_i^{m+2} - \cdots - a_m \sum x_i^{m+m} = 0 \end{cases}$$

$$\Rightarrow \begin{bmatrix} n & \sum x_i & \sum x_i^2 & \cdots & \sum x_i^m \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \cdots & \sum x_i^{m+1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \sum x_i^m & \sum x_i^{m+1} & \sum x_i^{m+2} & \cdots & \sum x_i^{2m} \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ \vdots \\ a_m \end{Bmatrix} = \begin{Bmatrix} \sum y_i \\ \sum x_i y_i \\ \vdots \\ \sum x_i^m y_i \end{Bmatrix}$$

مثال رگرسیون تک متغیره چندجمله ای

مثال/ نتایج بدست آمده از تصاویر MRI در مورد ارتباط بین زاویه خمش مفصل آرنج و طول بازوی گشتاور عضله در جدول داده شده است. با استفاده از رگرسیون چندجمله ای درجه ۲ برای θ ، طول بازوی گشتاور عضله را در زاویه ۳۵ درجه تخمین بزنید.

θ	0	20	40	60	80	100	120	140
$L(mm)$	13.4	24	30	40	48	52	49	44

رگرسیون چندجمله ای درجه ۲ برای θ

$$L = a_0 + a_1\theta + a_2\theta^2$$

$$\begin{bmatrix} n & \sum\theta & \sum\theta^2 \\ \sum\theta & \sum\theta^2 & \sum\theta^3 \\ \sum\theta^2 & \sum\theta^3 & \sum\theta^4 \end{bmatrix} \begin{Bmatrix} a_0 \\ a_1 \\ a_2 \end{Bmatrix} = \begin{Bmatrix} \sum L \\ \sum L\theta \\ \sum L\theta^2 \end{Bmatrix}$$

$$\sum\theta = 5.6000 \times 10^2$$
$$\sum\theta^2 = 5.6000 \times 10^4$$
$$\sum\theta^3 = 6.2720 \times 10^6$$
$$\sum\theta^4 = 7.4816 \times 10^8$$

$$\sum L = 3.0040 \times 10^2$$
$$\sum L\theta = 2.516 \times 10^4$$
$$\sum L\theta^2 = 2.5968 \times 10^6$$

مثال رگرسیون تک متغیره چندجمله ای

از حل ماتریس

$$a_0 = 11.533$$
$$a_1 = 0.68595$$
$$a_2 = -3.1429 \times 10^{-3}$$

رابطه رگرسیون درجه۲

$$L = -3.1429 \times 10^{-3}\theta^2 + 0.685595\theta + 11.533$$

$$\theta = 35 \rightarrow L = 31.6912mm$$

$$S_r = \sum e_i^2 = 38.756$$
$$S_t = \sum (L_i - \bar{L})^2 = 1320.5$$

$$r^2 = \frac{S_t - S_r}{S_t} = 0.97065$$

نکته آخر: در تمام رگرسیون های در نظر گرفته شده در این بخش، مدلها نسبت به ضرایب خطی بودند به همین دلیل به دستگاه معادلات خطی میرسیدیم. اگر رگرسیون غیرخطی نسبت به ضرایب باشند، به دستگاه معادله غیرخطی میرسیم.