LINEAR REGRESSION

Boi canh:

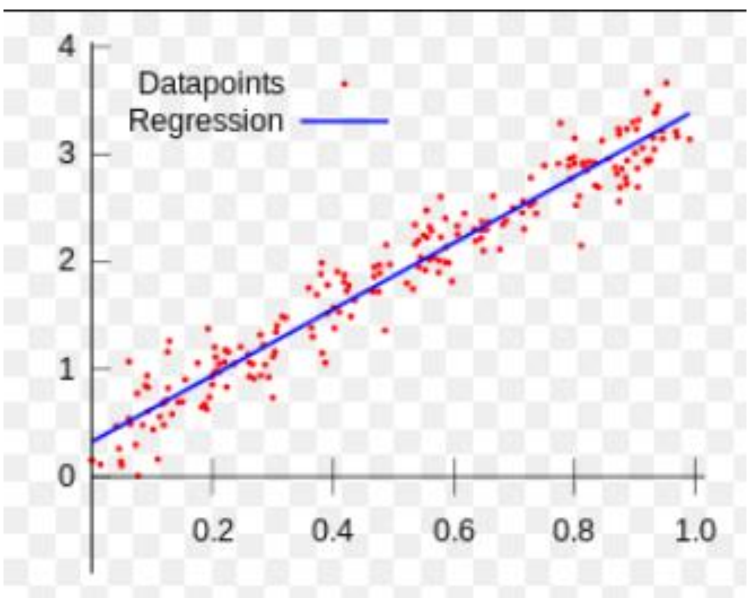
Cho mot tap hop cac diem du lieu: input x,y va output z

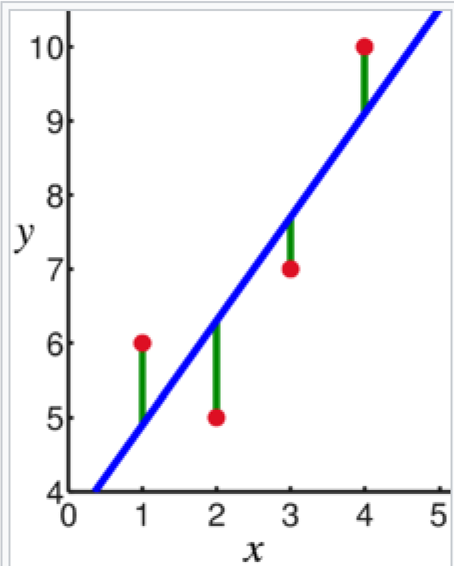
x1	y1	z1
x2	y2	z2
…	…	…
xk	yk	?

Vay neu co xk, yk, co the suy ra duoc zk?

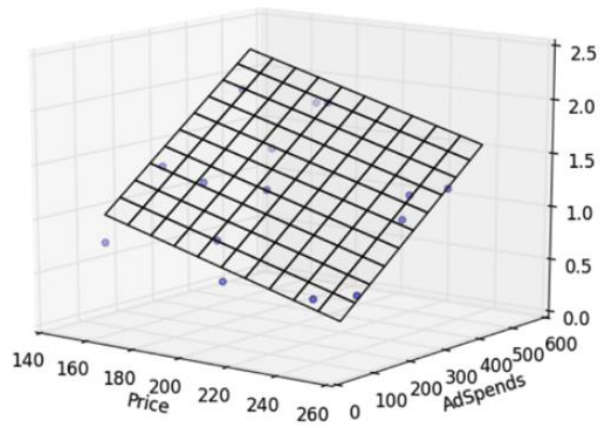
Tiep can:

Ta thay neu cac diem du lieu san co: p1(x1,y1,z1), p2(x1,y1,z1) … gan nhu phan bo tuyen tinh va thi ta co the ve mot mat phang p gan cac diem du lieu do nhat de su dung lam co so du doan cac diem du lieu tiep theo:





In linear regression, the observations (red) are assumed to be the result of random deviations (green) from an underlying relationship (blue) between the dependent variable (*y*) and independent variable (*x*).



Co the bieu dien p nhu sau: z = f(x, y) = w1.x1 + w2.x2 + 1.w3 ((w1,w2): coef, w3:intercept/bias)

Vay can tim (w1, w2, w3) de mat phang p la gan voi cac diem p1, p2 … pn nhat. Hay noi cach khac, chinh la tim phuong trinh mat phang p de tong cac sai so :

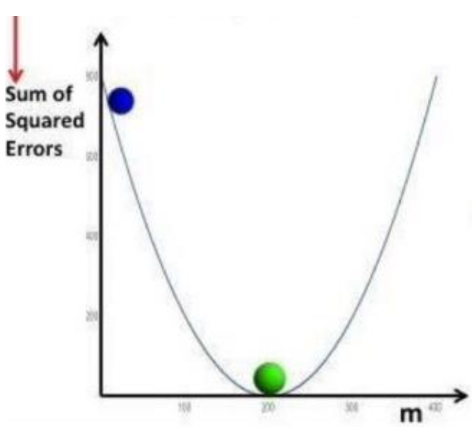
$$\text{£}(w) = \sum_{k=0}^{n} \frac{1}{2N}\left(z_i - (w_1.x_i + w_2.y_i + w_3)\right)^2$$

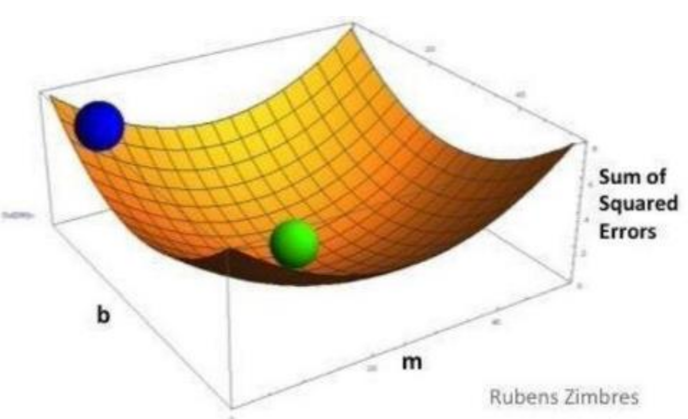
la nho nhat.

Dat **y** = $\begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$, Dat **X** = $\begin{bmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ & \vdots & \\ x_n & y_n & 1 \end{bmatrix}$, Dat **w** = $\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$

Ta co: $\text{£}(\boldsymbol{w}) = \frac{1}{2N}\|\boldsymbol{y} - \boldsymbol{X}.\boldsymbol{w}\|_2^2$

Nhan xet: de don gian hoa va bieu dien duoc duoi dang hinh hoc ta xet ham so z = £(w) voi w la vector 2 chieu (x, y) dang mot luoi parapol huong len.





Vay ham £(w) dat gia tri cuc tieu tai diem w* co dao ham £'(w) = 0. (*)

Ap dung (*) cho khong gian w € R^n

Bảng 2.1: Bảng các đạo hàm cơ bản.

$f(\mathbf{x})$	$\nabla f(\mathbf{x})$	$f(\mathbf{X})$	$\nabla_{\mathbf{X}} f(\mathbf{X})$
$\mathbf{x}$	$\mathbf{I}$	$\text{trace}(\mathbf{X})$	$\mathbf{I}$
$\mathbf{a}^T\mathbf{x}$	$\mathbf{a}$	$\text{trace}(\mathbf{A}^T\mathbf{X})$	$\mathbf{A}$
$\mathbf{x}^T\mathbf{A}\mathbf{x}$	$(\mathbf{A}+\mathbf{A}^T)\mathbf{x}$	$\text{trace}(\mathbf{X}^T\mathbf{A}\mathbf{X})$	$(\mathbf{A}+\mathbf{A}^T)\mathbf{X}$
$\mathbf{x}^T\mathbf{x} = \|\mathbf{x}\|_2^2$	$2\mathbf{x}$	$\text{trace}(\mathbf{X}^T\mathbf{X}) = \|\mathbf{X}\|_F^2$	$2\mathbf{X}$
$\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2$	$2\mathbf{A}^T(\mathbf{A}\mathbf{x}-\mathbf{b})$	$\|\mathbf{A}\mathbf{X}-\mathbf{B}\|_F^2$	$2\mathbf{A}^T(\mathbf{A}\mathbf{X}-\mathbf{B})$
$\mathbf{a}^T\mathbf{x}^T\mathbf{x}\mathbf{b}$	$2\mathbf{a}^T\mathbf{b}\mathbf{x}$	$\mathbf{a}^T\mathbf{X}\mathbf{b}$	$\mathbf{a}\mathbf{b}^T$
$\mathbf{a}^T\mathbf{x}\mathbf{x}^T\mathbf{b}$	$(\mathbf{a}\mathbf{b}^T+\mathbf{b}\mathbf{a}^T)\mathbf{x}$	$\text{trace}(\mathbf{A}^T\mathbf{X}\mathbf{B})$	$\mathbf{A}\mathbf{B}^T$

7.2.3 Nghiệm cho bài toán Linear Regression

Nhận thấy rằng hàm mất mát $\mathcal{L}(\mathbf{w})$ có đạo hàm tại mọi $\mathbf{w}$ (xem Bảng 2.1). Vậy việc tìm giá trị tối ưu của $\mathbf{w}$ có thể được thực hiện thông qua việc giải phương trình đạo hàm của $\mathcal{L}(\mathbf{w})$ theo $\mathbf{w}$ bằng không. Thật may mắn, đạo hàm của hàm mất mát của linear regression rất đơn giản:

$$\frac{\nabla\mathcal{L}(\mathbf{w})}{\nabla\mathbf{w}} = \frac{1}{N}\mathbf{X}(\mathbf{X}^T\mathbf{w}-\mathbf{y}) \tag{7.6}$$

Giải phương trình đạo hàm bằng không:

$$\frac{\nabla\mathcal{L}(\mathbf{w})}{\nabla\mathbf{w}} = 0 \Leftrightarrow \mathbf{X}\mathbf{X}^T\mathbf{w} = \mathbf{X}\mathbf{y} \tag{7.7}$$

Nếu ma trận $\mathbf{X}\mathbf{X}^T$ *khả nghịch*, phương trình (7.7) có nghiệm duy nhất $\mathbf{w} = (\mathbf{X}\mathbf{X}^T)^{-1}\mathbf{X}\mathbf{y}$.

Chu y: de phuong trinh (7.7) co nghiem thi (XX^T) phai kha nghich tuc la (XX^T) phai doc lap tuyen tinh. Gia su X thuoc R^{mxn} => A = XX^T thuoc R^{mxm} .

Vay dau tien ta phai di tim mot tap du lieu K thuoc X lon nhat co the de KK^T doc lap tuyen tinh. **Phuong phap**: Co the bat dau voi 1 tap dau tien 1 phan tu, Lap voi N diem du lieu con , them mot phan tu $\mathbf{x_i}$ moi vao tap du lieu **D** duoc tap du lieu moi $\mathbf{D_i}$ neu $\mathbf{D_i D_i}^T$ la phu thuoc tuyen tinh (det($\mathbf{D_i D_i}^T$) = 0) thi loai $\mathbf{x_i}$ ra.

Nếu phương trình:

$$\mathbf{0} = x_1\mathbf{a}_1 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n \tag{1.15}$$

có nghiệm duy nhất $x_1 = x_2 = \cdots = x_n = 0$, ta nói rằng hệ $\{\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n\}$ là một hệ *độc lập tuyến tính* (linear independence). Ngược lại, Nếu tồn tại $x_i \neq 0$ sao cho phương trình trên thoả mãn, ta nói rằng đó là một hệ *phụ thuộc tuyến tính* (linear dependence).

1.7.2 Tính chất

1. *Một hệ là phụ thuộc tuyến tính nếu và chỉ nếu tồn tại một vector trong hệ đó là tổ hợp tuyến tính của các vector còn lại.* Thật vậy, giả sử phương trình (1.15) có nghiệm khác không. Giả sử hệ số khác không là x_i, ta sẽ có:

$$\mathbf{a}_i = \frac{-x_1}{x_i}\mathbf{a}_1 + \cdots + \frac{-x_{i-1}}{x_i}\mathbf{a}_{i-1} + \frac{-x_{i+1}}{x_i}\mathbf{a}_{i+1} + \ldots \frac{-x_n}{x_i}\mathbf{a}_n \tag{1.16}$$

   tức $\mathbf{a}_i$ là một tổ hợp tuyến tính của các vector còn lại.