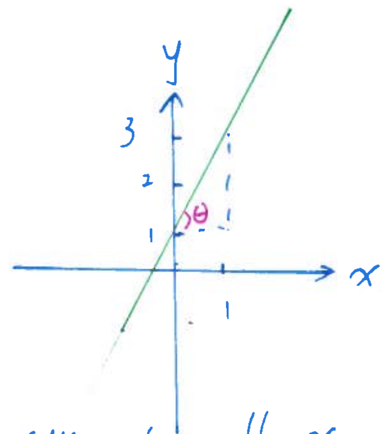
梯度下降

Slope

1元方程

$$y = 2x + 1$$

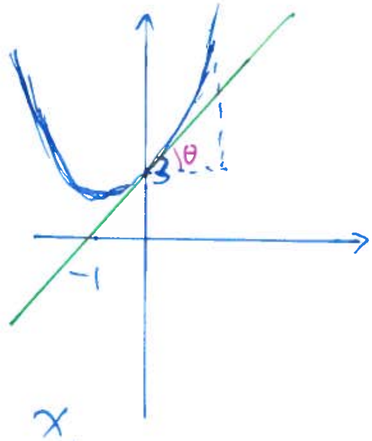


$$\frac{dy}{dx} = 2 = \tan\theta$$

$\frac{dy}{dx}$ is constant, slope it the same for all x

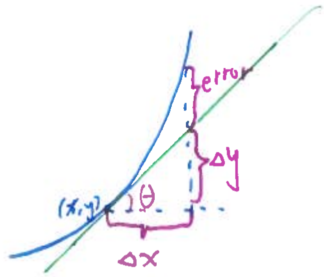
1元2次

$$y = (x+1)^2 + 2$$
$$= x^2 + 2x + 3$$



$$\frac{dy}{dx} = 2x + 2$$

$\frac{dy}{dx}$ is not the same for different x.



slope is like trend.

when $x = x + \Delta x$

then $y = y + \Delta y$

when $\Delta x \to 0$

then error $\to 0$

summary:

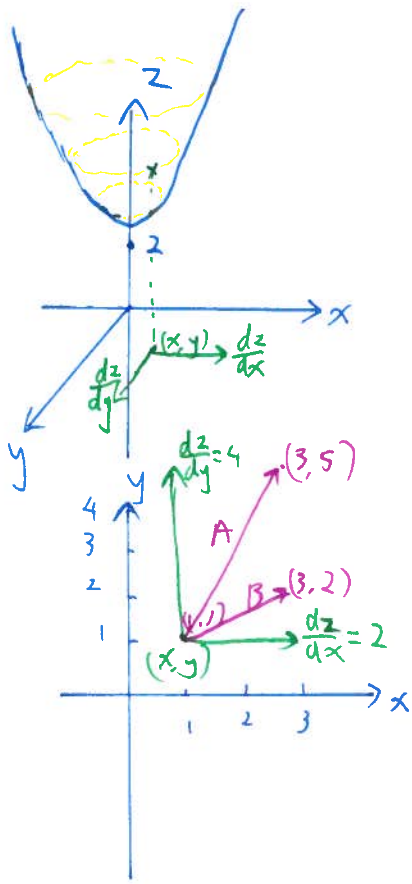
For a function $y=f(x)$. At point x, slope $\tan\theta = \frac{dy}{dx}$

when $x = x + 1$单位

then $y = y + \frac{dy}{dx} \cdot 1$单位

2 元

$$Z = x^2 + 2y^2 + 2$$



slope: $\begin{bmatrix} \frac{dz}{dx} \\ \frac{dz}{dy} \end{bmatrix} = \begin{bmatrix} 2x \\ 4y \end{bmatrix}$

At point $(x, y) = (1, 1)$

$$\begin{bmatrix} \frac{dz}{dx} \\ \frac{dz}{dy} \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix}$$

when: $x = x + 1$単位

$z = z + \frac{dz}{dx} \cdot 1$単位

when: $y = y + 1$単位

$z = z + \frac{dz}{dy} \cdot 1$単位

Increase in Z, if (1, 1) to (3, 5)

(1, 1) to (3, 2)

(1, 1) to (3, 5): $\Delta x = 3 - 1 = 2$

$\Delta y = 5 - 1 = 4$

$x = x + 2 \cdot$単位, $y = y + 4 \cdot$単位

$Z_x = Z + \frac{dz}{dx} \cdot 2$単位 $= Z + 4$単位

$Z_y = Z + \frac{dz}{dy} \cdot 4$単位 $= Z + 16$単位

(1, 1) to (3, 2): $\Delta x = 3 - 1 = 2$

$\Delta y = 2 - 1 = 1$

$x = x + 2 \cdot$単位

$y = y + 1 \cdot$単位

$Z_x = Z + \frac{dz}{dx} \cdot 2$単位 $= Z + 4$単位

$Z_y = Z + \frac{dz}{dy} \cdot 1$単位 $= Z + 4$単位

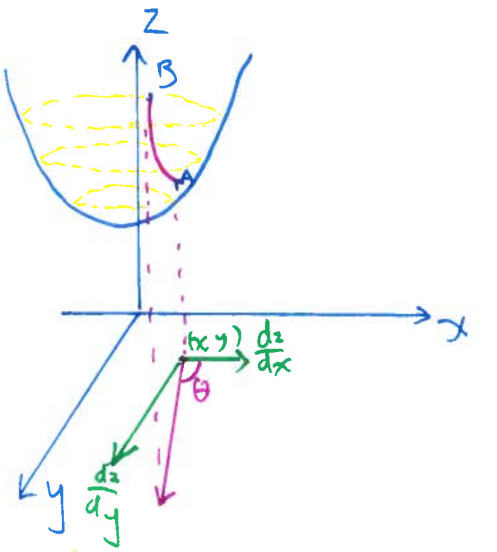
(1, 1) to (3, 5): $Z = Z_x + Z_y = 20$単位

(1, 1) to (3, 2): $Z = Z_x + Z_y = 8$単位

A, B both on $Z = x^2 + 2y^2 + 2$

If move from A to B, angle is θ

what is the change in Z ?

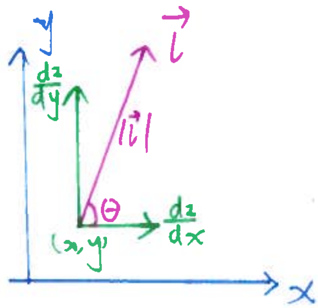


Assume vector from A to B is $\vec{l}$

Then, length of $\vec{l}$ is $|\vec{l}|$

Then $\quad x = x + |\vec{l}|\cos\theta$ 单位

$\qquad\ \ y = y + |\vec{l}|\sin\theta$ 单位

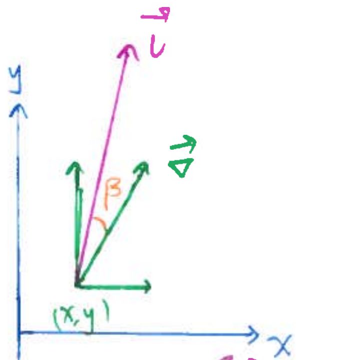


Therefore

$$Z = Z + \left[\frac{dz}{dx}\cdot|\vec{l}|\cos\theta + \frac{dz}{dy}\cdot|\vec{l}|\sin\theta\right] \text{单位}$$

$$= Z + \left[|\vec{l}|\cos\theta,\ |\vec{l}|\sin\theta\right]\cdot\begin{bmatrix}\frac{dz}{dx}\\ \frac{dz}{dy}\end{bmatrix}$$

$$= Z + \vec{l}\cdot\vec{\nabla} \text{ 单位}$$

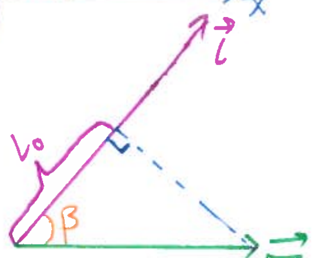


向量内积的几何意义

$\vec{l}\cdot\vec{\nabla} = |\vec{l}|\cdot|\vec{\nabla}|\cdot\cos\beta$

$= L_0$

$=$ $\vec{\nabla}$ 向量在 $\vec{l}$ 向量上的投影

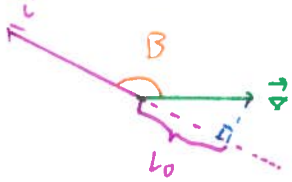


when $\beta = 0^\circ$, $\vec{l}\cdot\vec{\nabla}$ 值最大, Z 上升的最快

$\beta = 180^\circ$, $\vec{l}\cdot\vec{\nabla}$ 值最大(但是方向是反的), Z 下降最快

So, when $\vec{l} = \vec{\nabla}$, $\vec{l}$ 与 $\vec{\nabla}$ 方向一样, Z 上升最快

$\vec{l} = -\vec{\nabla}$, $\vec{l}$ 与 $\vec{\nabla}$ 方向相反, Z 下降最快,



要最小化 SSE, 就要选 $\vec{l} = -\vec{\nabla}$, 让 SSE 下降最快

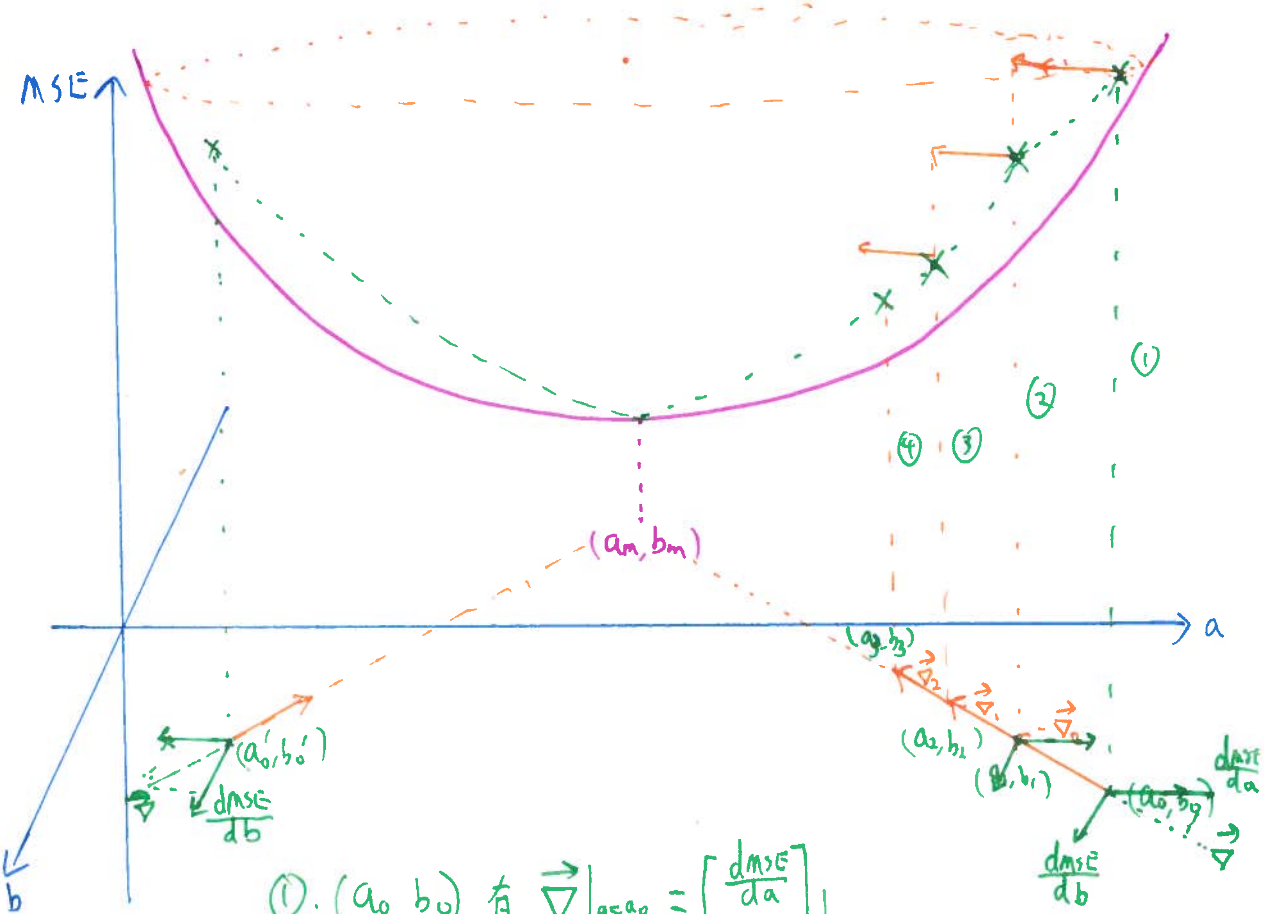
Review 1元线性回归

$y = ax + b$

$SSE = \sum_{i=1}^{n} (y_i - \hat{a} x_i - \hat{b})^2 \qquad MSE = \frac{1}{n} SSE$

梯度下降：

- 找一个起始点 (a_0, b_0)，算出在这个点的 $\vec{\nabla} = \begin{bmatrix} \frac{dMSE}{da} \\ \frac{dMSE}{db} \end{bmatrix} \Bigg|_{\substack{a=a_0 \\ b=b_0}}$
- 不断更新这个点 $\begin{bmatrix} a_1 \\ b_1 \end{bmatrix} = \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} - \underbrace{\lambda}_{\text{步长}} \cdot \vec{\nabla} \Big|_{\substack{a=a_0 \\ b=b_0}}$



①：(a_0, b_0)，有 $\vec{\nabla} \Big|_{\substack{a=a_0 \\ b=b_0}} = \begin{bmatrix} \frac{dMSE}{da} \\ \frac{dMSE}{db} \end{bmatrix} \Bigg|_{\substack{a=a_0 \\ b=b_0}}$

在 (a_0, b_0)，往 $\vec{\nabla} \Big|_{\substack{a=a_0 \\ b=b_0}}$ 方向，MSE 上升最快。

往 $-\vec{\nabla} \Big|_{\substack{a=a_0 \\ b=b_0}}$ 方向，MSE 下降最快。

更新 $\begin{bmatrix} a_0 \\ b_0 \end{bmatrix}$ 至 $\begin{bmatrix} a_1 \\ b_1 \end{bmatrix}$，$\begin{bmatrix} a_1 \\ b_1 \end{bmatrix} = \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} - \lambda \vec{\nabla} \Big|_{\substack{a=a_0 \\ b=b_0}}$

② 更新 $\begin{bmatrix} a_1 \\ b_1 \end{bmatrix}$ 至 $\begin{bmatrix} a_2 \\ b_2 \end{bmatrix}$，$\begin{bmatrix} a_2 \\ b_2 \end{bmatrix} = \begin{bmatrix} a_1 \\ b_1 \end{bmatrix} - \lambda \vec{\nabla} \Big|_{\substack{a=a_1 \\ b=b_1}}$

⋮

多元最小2乘

$$MSE = \frac{1}{n}(Y - X\theta)^T(Y - X\theta)$$

$$= \frac{1}{n}\left[Y^TY - \theta^TX^TY - Y^TX\theta + \theta^TX^TX\theta\right]$$

$$\vec{\nabla} = \frac{dMSE}{d\theta} = \frac{1}{n}\left[-2X^TY + 2X^TX\theta\right]$$

$$= \frac{2}{n}\left[X^T \cdot (X\theta - Y)\right]$$

$$= \begin{bmatrix} \frac{d}{d\theta_0} MSE \\ \vdots \\ \frac{d}{d\theta_p} MSE \end{bmatrix}$$

① Take a initial point $\theta_0 = \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_p \end{bmatrix}$, and a $\lambda = 0.01$

② For $m = 10000$ steps

$$\theta_{new} = \theta_{pre} - \lambda \nabla MSE(\theta)\Big|_{\theta = \theta_{pre}}$$