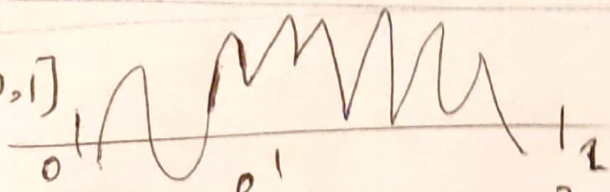
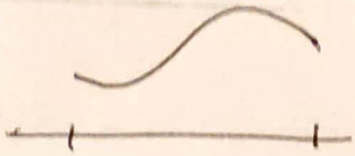
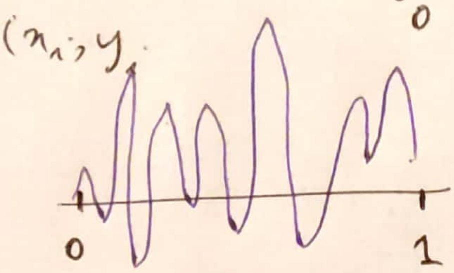
$p = 17 \qquad y = a_{17}x^{17} + a_{16}x^{16} + \cdots + a_2 x^2 + a_1 x + a_0 = f(x, \theta)$

$(x_i, y_i) \quad x_i \in [0,1]$

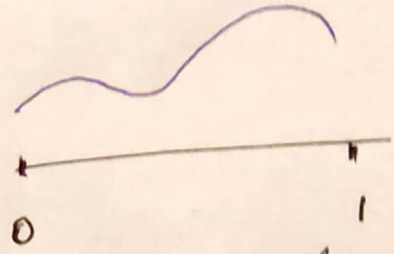




$$R(\theta) = R(a_0, a_1, \dots, a_p) = \int_0^1 \left(\frac{d}{dx} f(x,\theta)\right)^2 dx$$



R(θ) large



R(θ) small

$$\underline{R(\theta)} = \int_0^1 \left(\frac{d}{dx}\sum_{i=0}^{p} a_i x^i\right)^2 dx = \int_0^1 \left(\sum_{i=0}^{p} i a_i x^{i-1}\right)^2 dx$$

$$\Rightarrow R(\theta) = \theta^T M \theta$$

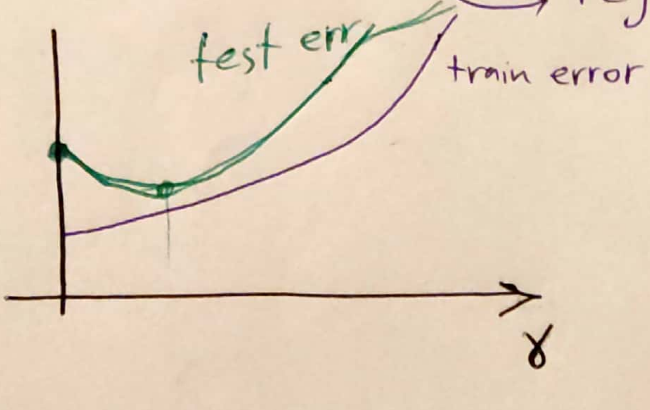
a simpler choice $\Leftarrow$ $R(\theta) = \theta^T\theta = \|\theta\|^2$

$$\min_\theta C(\theta)$$

$$\min_\theta C(\theta) + \gamma\, \underline{R(\theta)}$$

hyper parameter

regularizer



$C(\theta) = \|M\theta - y\|^2$ polynomial regression LA 32 (II)

$R(\theta) = \theta^T\theta = \|\theta\|^2$

$$\theta^* = \underset{\theta}{\operatorname{argmin}}\ C(\theta) + \gamma R(\theta)$$

$$= \underset{\theta}{\operatorname{argmin}}\ \|M\theta - y\|^2 + \gamma\|\theta\|^2$$

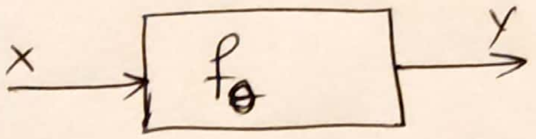
$$= \quad " \quad \|M\theta - y\|^2 + \|\sqrt{\gamma} I \theta\|^2$$

$$= \underset{\theta}{\operatorname{argmin}} \left\| \begin{bmatrix} M \\ \sqrt{\gamma} I \end{bmatrix} \theta - \begin{bmatrix} y \\ 0 \end{bmatrix} \right\|^2$$

$$\boxed{\|u\|^2 + \|v\|^2 = \left\|\begin{bmatrix} u \\ v \end{bmatrix}\right\|^2}$$

$$\|B\theta - z\|^2$$

$$\theta^* = (B^TB)^{-1}B^Tz$$



so far $f(x,\theta) = \phi(x)^T\theta$ linear in θ

$$f(x,\theta) = \tanh(x)\,a + (\sin x)b + (\log x)c + \exp(x)\,d$$

$$= [\tanh(x), \sin x, \log x, \exp(x)] \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

$$C(\theta) = \left\| \Phi \begin{bmatrix} \phi(x_1)^T \\ \phi(x_2)^T \\ \vdots \\ \phi(x_n)^T \end{bmatrix} \theta - \begin{bmatrix} y_1 \\ y_2 \\ | \\ y_n \end{bmatrix} \right\|^2$$

Scanned with CamScanner

LA32 (III)

$$C(\theta) = C\left(\begin{bmatrix}\theta_1\\ \theta_2\\ | \\ \theta_p\end{bmatrix}\right)$$

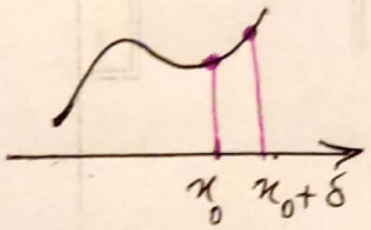
$$\left.\begin{aligned}
&\frac{\partial C}{\partial \theta_1}(\theta) = 0\\
&\frac{\partial C(\theta)}{\partial \theta_2} = 0\\
&\quad\vdots\\
&\frac{\partial C(\theta)}{\partial \theta_p} = 0
\end{aligned}\right\}$$

n equations

n unknowns $(\theta_1, \theta_2, \dots, \theta_p)$

solve for θ

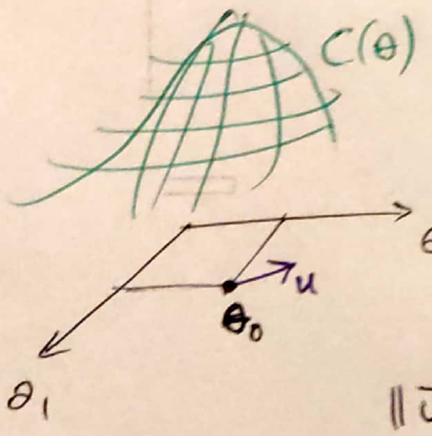
~~go~~ use gradient based optimization



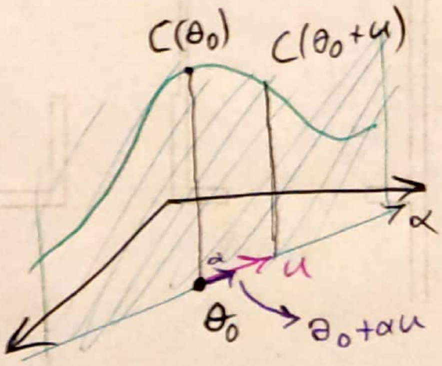
$$f'(x_0) = \left.\frac{df}{dx}\right|_{x=x_0} = ?$$

$$f'(x_0) = \lim_{\delta\to 0} \frac{f(x_0+\delta) - f(x_0)}{\delta}$$

$$C(\theta) = C(\theta_1, \theta_2, \dots, \theta_p)$$

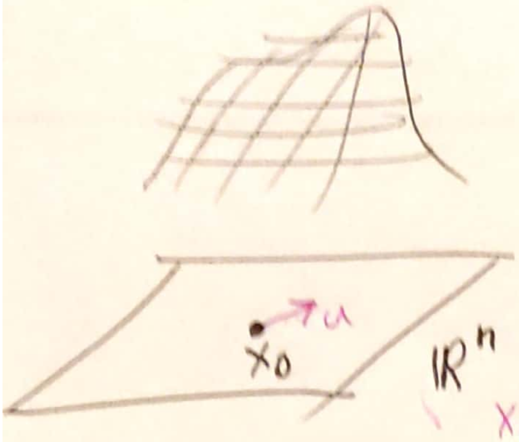


$\|\vec{u}\| = 1$



$g(\alpha) = C(\theta_0 + \alpha u)$

$$D[u]C(\theta_0) = \frac{C(\theta_0+\alpha u) - C(\theta_0)}{\alpha} = \left.\frac{d}{d\alpha} C(\theta_0+\alpha u)\right|_{\alpha=0}$$



$f(x)$

$f : \mathbb{R}^n \longrightarrow \mathbb{R}$

$$D[u]f(x_0) = \frac{d}{d\alpha} f(x_0 + \alpha u)\Big|_{\alpha=0}$$

$x_0 \in \mathbb{R}^n$, $u \in \mathbb{R}^n$

$\|u\| = 1$

Let u be any vector (not just a unit vector $\|u\| = 1$)

$$D[u]f(x_0) = \frac{d}{d\alpha} f(x_0 + \alpha u)\Big|_{\alpha=0} \quad \text{directional derivative}$$

$$D[\beta u]f(x_0) = \frac{d}{d\alpha} f(x_0 + \alpha \beta u)\Big|_{\alpha=0}$$

$$= \lim_{\alpha \to 0} \frac{\left(f(x_0 + \overbrace{\alpha\beta}^{\gamma} u) - f(x_0)\right)\beta}{\underbrace{\alpha\beta}_{\gamma}}$$

$$= \lim_{\gamma \to 0} \frac{f(x_0 + \gamma u) - f(x_0)}{\gamma}\, \beta$$

$$\Longrightarrow D[\beta u]f(x_0) = \beta\, D[u]f(x_0)$$

for differentiable functions

$$D[u+v]f(x_0) = D[u]f(x_0) + D[v]f(x_0)$$

$\Longrightarrow$ For ~~differ~~ a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$

$D[u]f(x_0) = D[u]f\big|_{x=x_0}$ is linear in $\vec{u}$.

$$D[u]f(x_0) = m^T u = \nabla^T u = \nabla(x_0)^T u$$

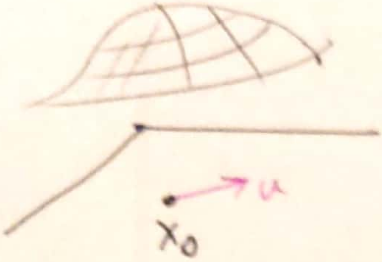
$D[\cdot]f\big|_{x_0} : \mathbb{R}^n \to \mathbb{R}$ $\quad u \in \mathbb{R}^n$ $\quad m \in \mathbb{R}^n$ $\quad \nabla$ gradient vector

LA32 IV

$D[u]f(x_0) = \nabla^T u = \langle \nabla, u \rangle \qquad \nabla \in \mathbb{R}^n$

$x_0 \in \mathbb{R}^n$

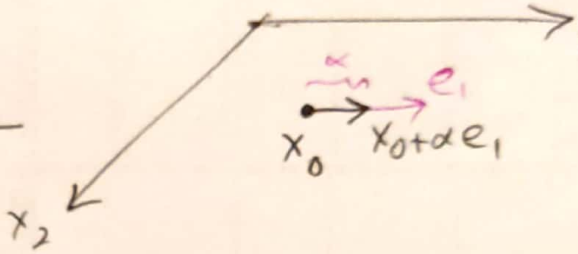
$u \in \mathbb{R}^n$



$$\nabla = \vec{m} = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix}$$

$$m_1 = \vec{m}^T \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \nabla^T \underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\vec{e}_1} = \nabla^T \vec{e}_1 = D[\vec{e}_1] f(x_0)$$

$$\lim_{\alpha \to 0} \frac{f(x_0 + \alpha e_1) - f(x_0)}{\alpha}$$



$$= \frac{\partial f}{\partial x_1} \Big|_{\vec{x} = \vec{x}_0}$$

$$\nabla = \begin{bmatrix} \partial f / \partial x_1 \\ \partial f / \partial x_2 \\ \vdots \\ \partial f / \partial x_n \end{bmatrix}$$

How to ~~comput~~ calculate ∇

راه حل اول: 1- find $\partial f / \partial x_1 \quad \frac{\partial f}{\partial x_2} \quad — \quad \frac{\partial f}{\partial x_n}$

2- arrange in a vector

$$\nabla = \begin{bmatrix} \partial f / \partial x_1 \\ \partial f / \partial x_2 \\ | \\ \partial f / \partial x_n \end{bmatrix}$$

Least squares

$$x^* = \arg\min \|Ax-b\|^2 \qquad x^* = (A^TA)^{-1}A^Tb$$

$$A = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \end{bmatrix} = \begin{bmatrix} r_1^T \\ r_2^T \\ \vdots \\ r_m^T \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & & & \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \qquad \|Ax-b\|^2 = \left\| \begin{bmatrix} r_1^T \\ r_2^T \\ \vdots \\ r_m^T \end{bmatrix} x - \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} \right\|^2$$

$$= \left\| \begin{bmatrix} r_1^Tx - b_1 \\ r_2^Tx - b_2 \\ \vdots \\ r_m^Tx - b_m \end{bmatrix} \right\|^2 = \sum_{i=1}^{m} (r_i^Tx - b_i)^2 = f(x)$$

$$= \sum_{i=1}^{m} (a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n - b_i)^2$$

$$\frac{\partial f}{\partial x_k} = \sum_{i=1}^{m} 2 \; a_{ik} \; (r_i^Tx - b)$$

$$= 2\sum_{i=1}^{m} a_{ik} (r_i^Tx - b)$$

$$= 2 \begin{bmatrix} a_{1k} & a_{2k} & \cdots & a_{mk} \end{bmatrix} \begin{bmatrix} r_1^Tx - b \\ r_2^Tx - b \\ \vdots \\ r_m^Tx - b \end{bmatrix}$$

$$= 2\, c_k^T \left(\begin{bmatrix} r_1^T \\ r_2^T \\ \vdots \\ r_m^T \end{bmatrix} x - b \right)$$

$$\frac{\partial f}{\partial x_k} = 2\, c_k^T (Ax-b)$$

$$\nabla = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \partial f / \partial x_2 \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = 2 \begin{bmatrix} c_1^T(Ax-b) \\ c_2^T(Ax-b) \\ \vdots \\ c_n^T(Ax-b) \end{bmatrix} = 2 \underbrace{\begin{bmatrix} c_1^T \\ c_2^T \\ \vdots \\ c_n^T \end{bmatrix}}_{A^T} (Ax-b) = \cancel{2A(Ax-b)}$$

$$= 2A^T(Ax-b)$$

VII

$$f(x) = \|Ax-b\|^2 \qquad A \in \mathbb{R}^{m\times n}$$

$$x^* = \underset{x}{\operatorname{argmin}}\, f(x)$$

$$\nabla_f(x) = 2A^T(Ax-b) = 2\underbrace{A^T}_{n\times m}\underbrace{(\underbrace{\underbrace{A}_{m\times n}\underbrace{x}_{n\times 1}-\underbrace{b}_{m\times 1}}_{m\times 1})}_{n\times 1 \,\in \mathbb{R}^n}$$

$$\nabla_f(x) = \vec{0} \Rightarrow 2A^T(Ax-b) = 0$$

$$\Rightarrow A^TAx - A^Tb = 0$$

$$\Rightarrow A^TAx = A^Tb \Rightarrow x^* = (A^TA)^{-1}A^Tb$$

$\cancel{f(x)}$ راه حل دوم محاسبه گرادیان

1- Calculate the diretional derivative

$$D[u]f(x_0) = \frac{d}{d\alpha} f(x+\alpha u)\Big|_{\alpha=0} = g(x,u)$$

2- Write $D[u]f$ in the form of $\nabla^T u$

$\langle \nabla, u\rangle$ ← گرادیان

$$f(x) = \|Ax-b\|^2 = (Ax-b)^T(Ax-b)$$

$$D[u]f(x) = \frac{d}{d\alpha} f(x+\alpha u)\Big|_{\alpha=0} = \cancel{\frac{d}{d\alpha}(A(x+\alpha u)-b)^T}$$

$$= \frac{d}{d\alpha}\left(A(\vec{x}+\alpha\vec{u})-b\right)^T\left(A(\vec{x}+\alpha\vec{u})-b\right)\Big|_{\alpha=0}$$

$$= (Au)^T(A(x+\alpha u)-b) + (A(x+\alpha u)-b)^TAu\Big|_{\alpha=0}$$

$$\overset{\alpha=0}{=} (Au)^T(Ax-b) + (Ax-b)^TAu$$

$$= 2(Ax-b)^TAu = \langle \underbrace{2A^T(Ax-b)}_{\nabla}, u\rangle$$

$$\Rightarrow \nabla = 2A^T(Ax-b)$$

$$D[u]f(x) = \underbrace{2(Ax-b)^T A}_{\nabla^T} u = \left(2A^T(Ax-b)\right)^T u$$

$$= \langle \nabla, u \rangle = \nabla^T u$$

$$\Rightarrow \boxed{\nabla = 2A^T(Ax-b)}$$

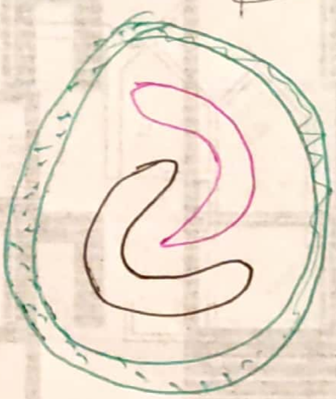
Spectral Clustering

k-means

Cluster

Clusteng

x_1 —— x_n



$\underset{n \times n}{A}$: similarity matrix

Affinity

$A^T = A$

$$D = \begin{bmatrix} \Sigma a & & 0 \\ & 0 & \\ 0 & & 0 \;\; 0 \end{bmatrix}$$

$L = D - A =$

$\boxed{\text{Spectral Clusteng}}$