$$\nabla_{\vec{\beta}} L \Big|_{\vec{\beta}^*} = 0 \qquad X^T \cdot \vec{y} = (X^T X) \vec{\beta}^*$$

$$X = \left[\vec{1} \,\vdots\, \vec{x}_1 \right]_{N \times 2} \qquad \text{invertible}$$

$$\vec{\beta}^* = \Big[\underset{2\times 2}{(X^T X)^{-1}} \underset{2\times N}{X^T} \Big] \vec{y}_{N \times 1}$$

(ii) $$H(L)_{ij} = \left(\frac{\partial^2 L}{\partial \beta_i \partial \beta_j} \right)$$

H is positive definite.

Aside::

$$\vec{\beta}_{(p+1)\times 1}$$

$$\vec{\nabla}_{\vec{\beta}} \Big(\underset{m\times 1}{\vec{f}(\vec{\beta})} \Big) = \Big(\vec{\nabla}_{\vec{\beta}}(f_1) \ \cdots \ \vec{\nabla}_{\vec{\beta}}(f_m) \Big)_{(p+1)\times m}$$

$$\vec{f} = \begin{pmatrix} f_1(\vec{\beta}) \\ \vdots \\ f_m \end{pmatrix}$$

$$H(L) = \vec{\nabla}_{\vec{\beta}} \left(\vec{\nabla}_{\vec{\beta}} L \right)$$

$$L = \frac{1}{N} \left[(\vec{y} - X\vec{\beta})^T (\vec{y} - X\vec{\beta}) \right]$$

$$\vec{\nabla}_{\vec{\beta}} L = \frac{-2}{N} \left[X^T \vec{y} - (X^T X) \vec{\beta} \right]$$

$$H = +\frac{2}{N} \vec{\nabla}_{\vec{\beta}} \left((X^T X) \vec{\beta} \right)$$

Aside:

$$\vec{\nabla}_{\vec{\beta}} (A\vec{\beta}) = \vec{\nabla}_{\vec{\beta}} \begin{bmatrix} \vec{a}_1^T \cdot \vec{\beta} \\ \vdots \\ \vec{a}_m^T \cdot \vec{\beta} \end{bmatrix} = \left[\vec{\nabla}_{\vec{\beta}} (\vec{a}_1^T \cdot \vec{\beta}) \ldots \vec{\nabla}_{\vec{\beta}} (\vec{a}_m^T \cdot \vec{\beta}) \right]$$

1

$$A = \begin{bmatrix} \vec{a}_1^T \\ \vdots \\ \vec{a}_m^T \end{bmatrix}$$

Reminder: $\vec{\nabla}_{\vec{\beta}} (\vec{\alpha}^T \cdot \vec{\beta}) = \vec{\alpha}$

1 $\vec{\nabla}_{\vec{\beta}} (A\vec{\beta}) = \left[\vec{a}_1 \; \ldots \; \vec{a}_m \right] = A^T$

Now back to H:

$$H = \frac{2}{N} \vec{\nabla}_{\vec{\beta}} \left((X^T X) \vec{\beta} \right) = \frac{2}{N} (X^T X)^T = \frac{2}{N} (X^T X)$$

Def: H is positive definite iff

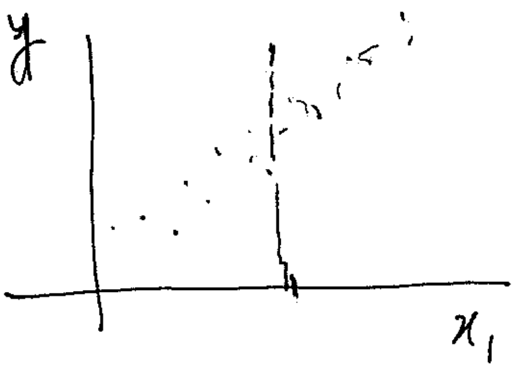
$$\vec{z}^T H \vec{z} > 0 \qquad \forall \vec{z} \neq 0$$

Show that $X^T X$ is positive definite:

$$\vec{z}^T X^T X \vec{z} = (X\vec{z})^T (X\vec{z}) = \|X\vec{z}\|^2 > 0$$
if $\vec{z} \neq 0$.

MSE is a convex function in the space of parameters

Statistical interpretation :



$p = 1$

$i = 1, \ldots N \quad (x_1^{(i)}, y^{(i)})$

$$y^{(i)} = \beta_0 + \beta_1 x_1^{(i)} + \varepsilon^{(i)}$$

$\varepsilon^{(i)}$ are i.i.d. from

Normal distribution with zero mean and variance σ^2 $\longleftarrow$ $\varepsilon \sim \mathcal{N}(0, \sigma^2)$

$$\forall i \quad \mathrm{Lik}(y^{(i)} \mid \vec{\beta}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\left(y^{(i)} - (\beta_0 + \beta_1 x_1^{(i)})\right)^2}{2\sigma^2}}$$

$$\mathrm{Lik}_{tot}(\vec{y} \mid \vec{\beta}) = \prod_{i=1}^{N} \mathrm{Lik}(y^{(i)} \mid \vec{\beta})$$

$$\mathcal{L}_{tot} = \sum_{i=1}^{N} \log\left(\mathrm{Lik}(y^{(i)} \mid \vec{\beta})\right) =$$

$$= C_1 - \frac{1}{2\sigma^2} \sum_{i=1}^{N} \left(\underbrace{y^{(i)} - (\beta_0 + \beta_1 x_1^{(i)}}_{e^{(i)}}\right)^2$$

$$= C_1 - \frac{1}{2\sigma^2} \vec{e}^{\,T} \cdot \vec{e} = C_1 - \frac{1}{2\sigma^2} N L_{MSE}$$

$$\mathcal{L}_{tot} = C_1 - \frac{N}{2\sigma^2} L_{MSE}$$

$$-\frac{\partial \mathcal{L}_{tot}}{d\vec{\beta}} \longleftrightarrow \frac{d L_{MSE}}{d\vec{\beta}}$$

maximum likelihood

minimal loss (MSE)

Given $\{(x_1^{(i)}, y^{(i)})\}_{i=1}^{N}$

the linear model $\therefore f(x_1) = (1 \;\; x)\begin{pmatrix}\beta_0 \\ \beta_1\end{pmatrix}$

$$= (1 \;\; x)\cdot\vec{\beta}$$

minimises the loss

$$L = \frac{1}{N}\left[(\vec{y} - X\vec{\beta})^T(\vec{y} - X\vec{\beta})\right]$$

when $\vec{\beta}^* = \underbrace{(X^T X)^{-1} X^T}_{X^+}\vec{y}$

Moore-Penrose pseudoinverse of X

$p=1$

$$\underset{N\times 2}{X}\,\vec{\beta} = \underset{N\times 1}{\vec{y}}$$

No solution! (overdetermined)

$$\downarrow$$

$$X^T X \vec{\beta} = X^T \vec{y}$$

$$\vec{\beta} = (X^T X)^{-1} X^T \vec{y}$$

Some properties of X^+:

Equivalent properties for invertible A:

$$X^+ X X^+ = X^+ \longleftrightarrow (A^{-1} A A^{-1} = A^{-1})$$

$$X X^+ X = X \longleftrightarrow (A A^{-1} A = A)$$

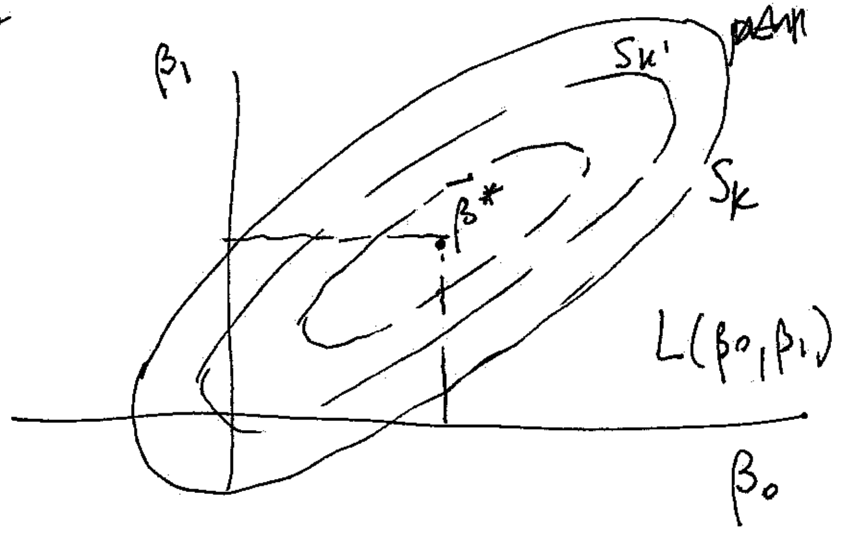
Optimisation $\equiv$ minimised

$$L(\vec{\beta}) = \frac{1}{N}\left((\vec{y} - X\vec{\beta})^T (\vec{y} - X\vec{\beta})\right)$$

$$\vec{\nabla}_{\vec{\beta}} L \Big|_{\vec{\beta}^*} = 0 \quad (*)$$

We will not always be able to solve $(*)$ for the normal equations.

$p = 1$



$\vec{\beta} = (\beta_0, \beta_1)$

$k' < k$

$$S_k = \{ \vec{\beta} \mid L(\vec{\beta}) = k \}$$