Regression

8/24

Recap:

$$\underbrace{y_i}_{\text{data}} = \underbrace{\beta_0 + \beta_1 x_i}_{\text{signal}} + \underbrace{\varepsilon_i}_{\text{noise}}$$

Using the LS Principle we tried to find estimates $(\hat{\beta}_0, \hat{\beta}_1)$ of (β_0, β_1).

Specifically, the LS Estimators are

$$\underline{(\hat{\beta}_0, \hat{\beta}_1)} = \underset{\beta_0, \beta_1}{\operatorname{argmin}}\ Q(\beta_0, \beta_1)$$

$$= \underset{\beta_0, \beta_1}{\operatorname{argmin}} \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

We found the closed form:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

Are these estimators any good?
What properties do they have?

<u>Gauss-Markov:</u>

Under the SLR model assumptions, the LS estimators $\hat{\beta}_0$ & $\hat{\beta}_1$ are

i. unbiased for β_0 & β_1 resp.

& ii. they are BLUE.

"Best Linear Unbiased Estimators"

Focus on $\hat{\beta}_1$:

How can I show that $\hat{\beta}_1$ is unbiased for β_1?

unbiased $\Leftrightarrow$ $\text{bias}(\hat{\beta}_1)=0 \Leftrightarrow \underline{E(\hat{\beta}_1)=\beta_1}$

Want to show

$$E(\hat{\beta}_1) = E\left(\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2}\right)$$

Def: $SSXY = \sum_i (x_i-\bar{x})(y_i-\bar{y})$

$SSX = \sum_{i=1}^{n}(x_i-\bar{x})^2$

Lemma:

$$SSXY = \sum_i \underbrace{(x_i-\bar{x})(y_i-\bar{y})} = \sum_{i=1}^{n}(x_i-\bar{x})y_i$$

why?

$$= \sum_i [(x_i-\bar{x})y_i - (x_i-\bar{x})\bar{y}]$$

$$= \sum_i (x_i-\bar{x})y_i - \sum_i (x_i-\bar{x})\bar{y}$$

$$= \sum_i (x_i - \bar{x}) y_i - \bar{y}\left(\sum_i (x_i - \bar{x})\right)$$

$$\text{"} \qquad - \bar{y}\left(\sum_i x_i - n\bar{x}\right)$$

$$\text{"} \qquad - \bar{y}\left(n\bar{x} - n\bar{x}\right)$$

$$\text{"} \qquad - \bar{y}(0)$$

$$\bar{x} = \frac{1}{n} \sum_i x_i$$

$$n\bar{x} = \sum_i x_i$$

$$= \sum_i (x_i - \bar{x}) y_i$$

$$\underline{E(\hat{\beta}_1)} = E\left(\frac{SSXY}{SSX}\right) = E\left(\frac{\sum_{i=1}^n (x_i - \bar{x}) y_i}{SSX}\right)$$

Define $\boxed{k_i = \frac{x_i - \bar{x}}{SSX}}$

$$= E\left(\sum_{i=1}^n k_i y_i\right)$$

$$= \sum_{i=1}^{n} k_i E(y_i)$$

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$E(y_i) = E(\beta_0 + \beta_1 x_i + \varepsilon_i)$$

$$= \underbrace{\beta_0 + \beta_1 x_i + \cancel{E(\varepsilon_i)}}$$

$$= \sum_{i=1}^{n} k_i (\beta_0 + \beta_1 x_i)$$

$$= \beta_0 \cancel{\sum_{i=1}^{n} k_i}^{\,0} + \beta_1 \underbrace{\cancel{\sum_{i=1}^{n} k_i x_i}}^{\,1} = \beta_1 \quad \checkmark \quad \square$$

① $$\sum_{i=1}^{n} k_i = \sum_{i=1}^{n} \frac{x_i - \bar{x}}{SSX}$$

$$= \frac{1}{SSX} \underbrace{\left(\cancel{\sum_i (x_i - \bar{x})}\right)}^{\,0} = \frac{0}{SSX} = 0$$

$$SSX = \sum_{i=1}^{n} (x_i - \bar{x})^2 \geq 0$$

x_1	x_2	x_3
1	2	3

$\bar{x} = 2$

$$SSX = (\underbrace{1-2}_{i=1})^2 + (2-2)^2 + (3-2)^2$$

$$= 1 + 0 + 1 = 2$$

② $$\sum_{i=1}^{n} k_i x_i = \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{SSX}\right) x_i$$

$$= \frac{1}{SSX} \sum_{i=1}^{n} (x_i - \bar{x}) x_i = \frac{SSX}{SSX} = 1$$

$$\sum_{i=1}^{n} (x_i^2 - x_i \bar{x})$$

$$\Sigma_i x_i^2 - \bar{x} \Sigma_i x_i$$

$$\Sigma_i x_i^2 - n\bar{x}^2$$

$$\underline{SSX} = \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_i (x_i^2 - 2x_i\bar{x} + \bar{x}^2)$$

$$= \sum_i x_i^2 - 2n\bar{x}^2 + n\bar{x}^2$$

$$= \underline{\sum_i x_i^2 - n\bar{x}^2}$$

BLUE

"best" — $\boxed{Var(\hat{\beta}_1) \leq Var(\underline{\tilde{\beta}_1})}$

"linear" — $\hat{\beta}_1 = \sum_i k_i y_i$

"unbiased" — $E(\hat{\beta}_1) = \beta_1$

Min Variance Sketch

I know $\hat{\beta}_1 = \sum_i k_i y_i$

Consider a different alternative linear estimator

$$\tilde{\hat{\beta}}_1 = \sum_i \tilde{k}_i y_i \qquad d_i = \tilde{k}_i - k_i$$

$$= \sum_i (\underset{=}{k_i} + \underset{\sim}{d_i}) y_i \quad \text{where } d_i \neq 0 \ \forall i$$

Can I show that

$$Var(\hat{\beta}_1) \leq Var(\tilde{\hat{\beta}}_1)$$

① ②

①

$$Var(\hat{\beta}_1) = Var\left(\sum_{i=1}^{n} k_i y_i\right)$$

by cov assump

$$= \sum_{i=1}^{n} Var(k_i y_i)$$

$$= \sum_{i=1}^{n} k_i^2 \, var(y_i)$$

$$= \sum_{i=1}^{n} k_i^2 \sigma^2 = \sigma^2 \sum_{i=1}^{n} k_i^2$$

$$= \frac{\sigma^2}{SSX}$$

$$= \frac{\sigma^2}{\sum_i (x_i - \bar{x})^2}$$

$$\sum_{i=1}^{n} k_i^2 = \sum_{i=1}^{n} \left[\frac{(x_i - \bar{x})}{SSX} \right]^2$$

$$= \sum_{i=1}^{n} \frac{(x_i - \bar{x})^2}{(SSX)^2}$$

$$= \frac{1}{(SSX)^2} \underbrace{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$= \frac{SSX}{(SSX)^2} = \frac{1}{SSX}$$

② $$Var(\tilde{\beta}_1) = Var(\textstyle\sum_i (k_i + d_i) y_i)$$

$$= \sum_{i=1}^{n} (k_i + d_i)^2 Var(y_i)$$

$$= \sigma^2 \sum_{i=1}^{n} (k_i + d_i)^2$$ (y: all shown)

$$= \sigma^2 \left(\sum_i k_i^2 + \cancel{2 \sum_i k_i d_i}^{\,0} + \sum_i d_i^2 \right)$$

$$= \underbrace{\sigma^2 \sum_i k_i^2}_{Var(\hat{\beta}_1)} + \underbrace{\sigma^2 \sum_i d_i^2}_{\geq 0}$$

$$Var(\tilde{\beta}_1) = Var(\hat{\beta}_1) + \underline{\sigma^2 \sum_i d_i^2} \geq Var(\hat{\beta}_1)$$

(N)

$$\hookrightarrow \hat{\beta}_0 \sim \cancel{N}\left(\beta_0, \sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{SSx}\right)\right)$$

$$\hat{\beta}_1 \sim N\left(\beta_1, \frac{\sigma^2}{\sum_i (x_i-\bar{x})^2}\right)$$

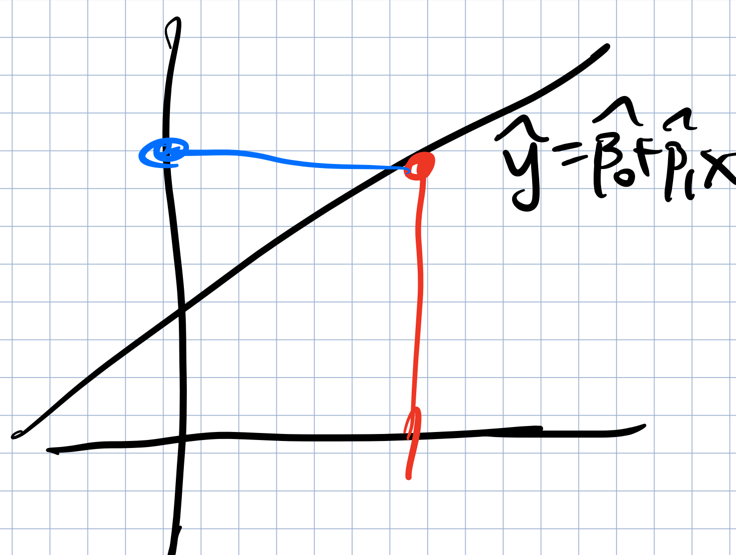
When doing prediction, we

least squares reg. line

can use the LSRL

to predict responses like this:

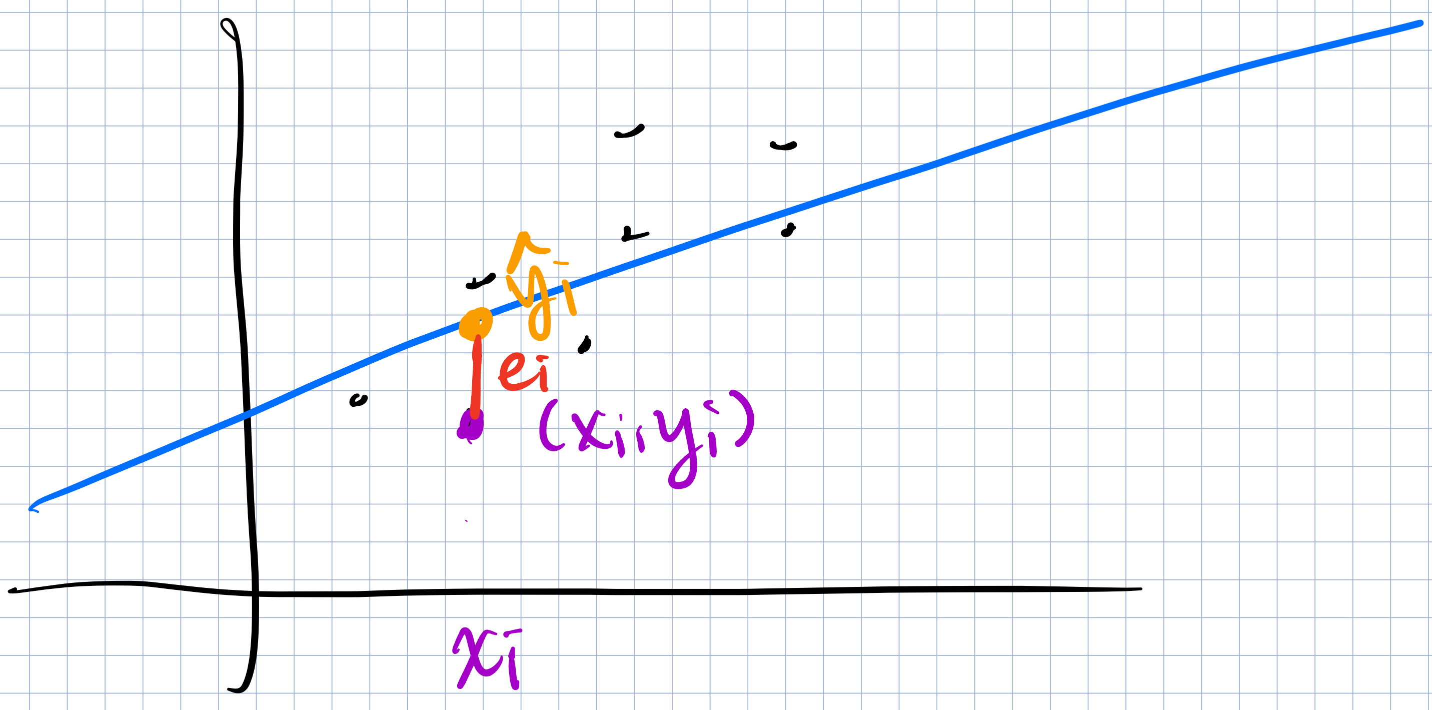
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$



Specifically if we're referring to a prediction for an in-sample obs. $X = x_i$:

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

This is called a "fitted value" of the i^{th} observation.



The residual

$$e_i := y_i - \hat{y}_i$$

and it aims to quantify prediction error for the i^{th} observation.

Difference b/w e_i & ε_i

$$\varepsilon_i = Y_i - \beta_0 - \beta_1 X_i$$

not

vs.

calculatable

$$e_i = y_i - \hat{y}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i$$

The residual provides a guess / a prediction of its corresponding theoretical error.

- $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$
- $e_i \sim N\left(\underline{\underline{0}}, \sigma^2\left[1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{SSX}\right]\right)$

$E(e_i) = 0$ (your HW)

$Var(e_i) = \sigma^2\left(1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{SSX}\right)$

Try to show:

$\sum_j h_j y_j$

$Var(e_i) = Var(y_i - \hat{y}_i)$

$= Var(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$

$= Var(\cancel{\beta_0 + \beta_1 x_i} + \varepsilon_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$

$$= Var(\varepsilon_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)$$

$$= \underline{Var(\varepsilon_i)}^{①} + \underline{Var(\hat{\beta}_0)}^{②} + \underline{Var(\hat{\beta}_1 x_i)}^{③}$$

$$+ 2Cov(\varepsilon_i, -\hat{\beta}_0) \quad ④$$

$$+ 2Cov(\varepsilon_i, -\hat{\beta}_1 x_i) \quad \underline{⑤}$$

$$+ 2Cov(\hat{\beta}_0, \hat{\beta}_1 x_i) \quad ⑥$$

$$\boxed{\begin{aligned} &Var(X+Y) = \\ &Var(X) + Var(Y) \\ &\quad + 2Cov(X,Y) \end{aligned}}$$

① $Var(\varepsilon_i) = \sigma^2$

② $Var(\hat{\beta}_0) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{SSX}\right)$

③ $Var(\hat{\beta}_1 x_i) = x_i^2 Var(\hat{\beta}_1)$

$$= \frac{x_i^2 \sigma^2}{SSX}$$

what is c_j^2.

④ $2Cov(\varepsilon_i, -\hat{\beta}_0)$

$$= -2Cov(\varepsilon_i, \sum_{j=1}^{n} c_j y_j)$$

What are c_js?

$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

$$= \frac{1}{n}\sum_{j=1}^{n} y_j - \left(\sum_{j=1}^{n} k_j y_j\right)\bar{x}$$

$$= \frac{1}{n}\sum_{j=1}^{n} y_j - \bar{x}\sum_{j=1}^{n} k_j y_j$$

$$= \sum_{j=1}^{n}\left(\frac{1}{n} y_j - \bar{x} k_j y_j\right)$$

$$= \sum_{j=1}^{n} \underbrace{\left(\frac{1}{n} - \bar{x} k_j\right)}_{\substack{\shortparallel \\ c_j}} y_j$$

$$= -2\,\mathrm{Cov}\left(\varepsilon_i, \sum_{j=1}^{n} c_j y_j\right)$$

$$= -2\sum_{j=1}^{n} \mathrm{Cov}(\varepsilon_i, c_j y_j)$$

$$= -2\,\mathrm{Cov}(\varepsilon_i, c_i y_i)$$

$$= -2 c_i\,\mathrm{Cov}(\varepsilon_i, \cancel{\beta_0 + \beta_1 X_i} + \varepsilon_i)$$

$$= -2 \underline{\underline{c_i}}\, \sigma^2$$

⑤

$$2\,\mathrm{Cov}(\varepsilon_i, -\hat{\beta}_1 x_i)$$

$$= -2x_i \,\mathrm{Cov}\Big(\varepsilon_i, \sum_{j=1}^{n} k_j y_j\Big)$$

$$= -2x_i \,\mathrm{Cov}(\varepsilon_i, \underline{k_i y_i})$$

$$= -2k_i x_i \,\mathrm{Cov}(\varepsilon_i, \varepsilon_i)$$

$$= -2\underline{\underline{k_i}} x_i \sigma^2$$

⑥

$$2\,\mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1 x_i)$$

$$= 2x_i \,\mathrm{Cov}\Big(\sum_{j=1}^{n} c_j y_j, \sum_{\ell=1}^{n} k_\ell y_\ell\Big)$$

$$= 2x_i \sum_{m=1}^{n} \mathrm{Cov}(c_m y_m, k_m y_m)$$

$$= 2 x_i \sum_{m=1}^{n} \underline{\underline{c_m k_m}} \, \sigma^2$$

$$\Longrightarrow$$

$$Var(e_i) = 1 + 2 + 3 + 4 + 5 + 6$$

$$= \sigma^2 \left(1 - \frac{1}{n} - \frac{(x_i - \bar{x})^2}{SSX} \right)$$

⊛ For full worked out version, see AM notes or Supp. Notes!

Last Time

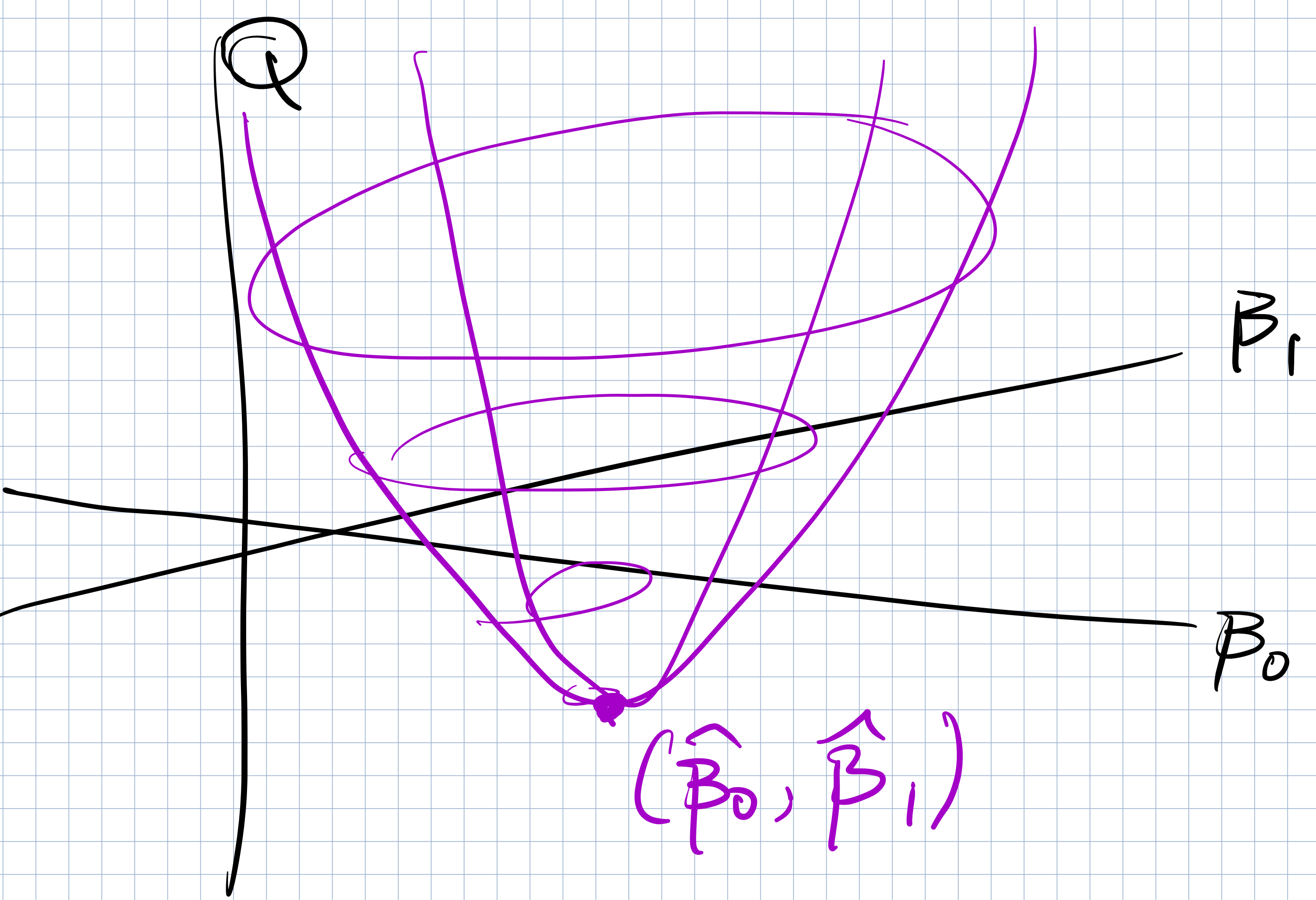
$$(\hat{\beta}_0, \hat{\beta}_1) = \underset{\beta_0, \beta_1}{\operatorname{argmin}}\, Q(\beta_0, \beta_1)$$

$$\Downarrow \quad \frac{\partial Q}{\partial \beta_0} \overset{!}{=} 0 \quad \& \quad \frac{\partial Q}{\partial \beta_1} \overset{!}{=} 0$$

To guarantee that the LS ests are actually minimizers, we have to look @ the Hessian matrix.

$$H = \begin{bmatrix} \frac{\partial^2 Q}{\partial \beta_0^2} & \frac{\partial^2 Q}{\partial \beta_0 \partial \beta_1} \\ \frac{\partial^2 Q}{\partial \beta_0 \partial \beta_1} & \frac{\partial^2 Q}{\partial \beta_1^2} \end{bmatrix}$$

$$= \begin{bmatrix} 2n & 2n\bar{x} \\ 2n\bar{x} & 2\sum_i x_i^2 \end{bmatrix}$$



We need to check that

H is positive def.

$\forall a \in \mathbb{R}^2$

H is p.d. $\Longleftrightarrow$ $\underline{a^T H a} > 0$

if $\underline{a \neq 0}$

$H = Q \Lambda Q^T$ $\quad$ $\mathrm{diag}(\Lambda)_j > 0$ $\quad \forall j = 1, 2$