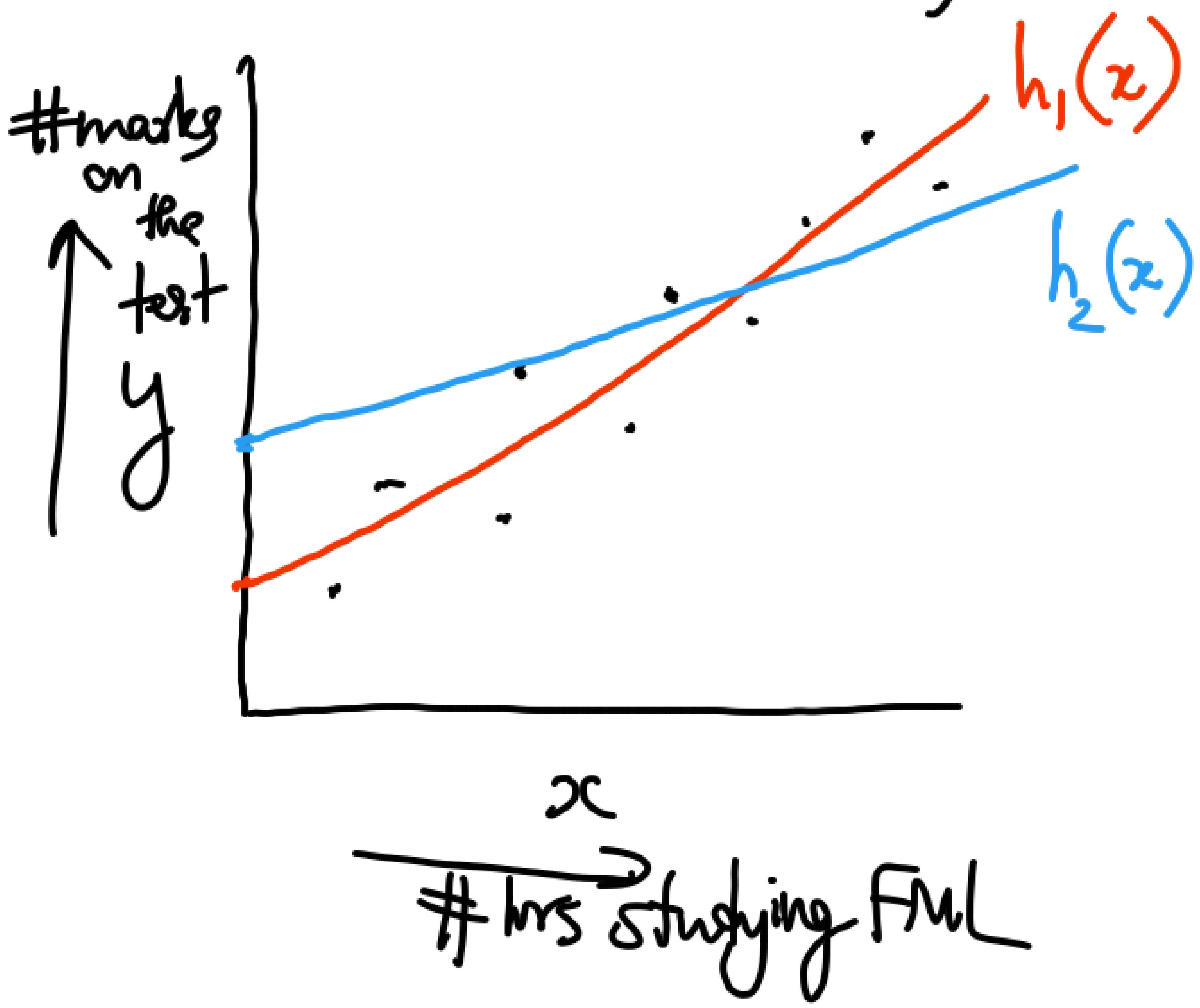
CS 725

Consider the following 1D case of linear regression



$$h_1(x) = 1 + 0.5x$$

$$h_2(x) = \underbrace{2}_{\text{intercept}} + \underbrace{0.2}_{\text{slope}}x$$

$$h_w(x) = \underbrace{w_0 + w_1}_{\text{parameters or weights}} x$$

$$h_w(x) = w^T \vec{x}$$

[VECTORIZED FORM OF LINEAR REGRESSION]

where $w = \begin{bmatrix} w_0 \\ w_1 \end{bmatrix}_{2\times 1}$, $\vec{x} = \begin{bmatrix} 1 \\ x \end{bmatrix}$

Easily extend to d-dimensional points:

$$h_w(x) = w^T \vec{x}, \text{ where } w = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix}, \vec{x} = \begin{bmatrix} 1 \\ x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$$

$(\vec{x} \in \mathbb{R}^{d+1})$

HYPOTHESIS SPACE FOR LINEAR REGRESSION

$$\mathcal{H} = \left\{ h_w : h_w(x) = w^T x, \ w \in \mathbb{R}^{d+1} \right\}$$

LINEAR IN LINEAR REGRESSION

Linear fn of its parameters

Q2) Measuring the performance of the predictor func

Loss function/error function : $\mathcal{L}(h, \mathcal{D}_{train}) \equiv \mathcal{L}(w, \mathcal{D}$

$$\mathcal{D}_{train} = \{(x_i, y_i)\}_{i=1}^{n}$$

$y_i -$

RESID

Good loss function for Linear Regression : LEAST SQUARES L

OR MEAN SQUARE LOS

defined as

$$\boxed{\mathcal{L}_{MSE}(w, \mathcal{D}) = \frac{1}{n}\sum_{i=1}^{n}\Big(y_i - \underbrace{h_w(x_i)}\Big)^2}$$

$\hat{y}_i$ (

ction 2

$\Theta))$

$\hat{y}_i$

UAL

LOSS

SS

Predicted value)

$$\mathcal{L}_{MSE}(w, \mathcal{D}) = \frac{1}{n} \sum_i \left(y_i - w^T x_i\right)^2$$

ANOTHER LOSS CANDIDATE: $$\mathcal{L}_{MAE}(w, \mathcal{D}) = \frac{1}{n} \sum_i \left| y_i - w^T x_i \right|$$

(MEAN ABSOLUTE ERROR)

L_2 norm $\|x\|_2 = \sqrt{x_1^2 + x_2^2 + \ldots + x_d^2}$

Q3) How do we find the "best" predictor function?

best $\mapsto$ the one that minimizes the loss function

OPTIMIZATION PROBLEM FOR LINEAR REGRESSION

$$w^* = \underset{w}{\operatorname{argmin}}\, L(w, \mathcal{D}) = \underset{w}{\operatorname{argmin}}\, \frac{1}{n} \sum_{i=1}^{n} (y_i - w^T x_i)^2$$

$$= \underset{w}{\operatorname{argmin}}\, \|\vec{y} - X\vec{w}\|_2^2$$

where $\vec{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$, $X = \begin{bmatrix} \leftarrow x_1^T \rightarrow \\ \vdots \\ \leftarrow x_n^T \rightarrow \end{bmatrix}_{n \times (d+1)}$, $\vec{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix}_{(d+1) \times 1}$

MSE loss is <u>CONVEX</u>

$$w^* = \underset{w}{\operatorname{argmin}} \sum_i \left(y_i - w^T x_i\right)^2$$

Solve for a simple 1D case:

$$w_0^*, w_1^* = \underset{w_0, w_1}{\operatorname{argmin}} \overbrace{\sum_i \left(y_i - w_0 - w_1 x_i\right)^2}^{\mathcal{L}(w, \mathcal{D})}$$

Compute the derivative of $\mathcal{L}$ w.r.t. w_0, w_1, equate to 0 and solve for them.

$$w_0^*, w_1^* = \underset{w_0, w_1}{\operatorname{argmin}} \overbrace{\sum_{i=1}^{n} \left(y_i - w_0 - w_1 x_i\right)^2}^{\mathcal{L}}$$

$$\frac{\partial \mathcal{L}}{\partial w_0} = 0 : \quad -2\sum_i \left(y_i - w_0 - w_1 x_i\right) = 0 \Rightarrow \sum_i y_i = n w_0 + w_1 \sum_i x_i$$

$$\frac{\partial \mathcal{L}}{\partial w_1} = 0 : \quad -2\sum_i \left(y_i - w_0 - w_1 x_i\right) x_i = 0 \qquad \Rightarrow \boxed{w_0 = \bar{y} - w_1 \bar{x}}$$

where $\bar{y} = \frac{\sum_i y_i}{n}$, $\bar{x} = \frac{\sum_i}{n}$

$$\Rightarrow \sum_i y_i x_i = \sum_i w_0 x_i + \sum_i w_1 x_i^2$$

$$\Rightarrow \boxed{w_1 = \frac{\frac{\sum_i x_i y_i}{n} - \bar{x}\bar{y}}{\frac{\sum_i x_i^2}{n} - \bar{x}^2}}$$

Extending to d-dimensional inputs

CLOSED FORM SOLUTION FOR LINEAR REGRESSION

$$W^* = \underset{w}{\operatorname{argmin}} \underbrace{\sum_i \left(y_i - w^T x_i\right)^2}_{L}$$

$$\nabla_w L = \begin{bmatrix} \frac{\partial L}{\partial w_1} \\ \frac{\partial L}{\partial w_2} \\ \vdots \\ \frac{\partial L}{\partial w_d} \end{bmatrix} \text{ ; Compute } \nabla_w L = 0, \text{ solve for } w \text{ ;}$$

$$-2 \sum_i \left(y_i - w^T x_i\right) x_i = 0$$

$$\Rightarrow \sum_i y_i x_i + \sum_i \left(w^T x_i\right) x_i = 0$$

$$-X^T y + X^T X w = 0$$

$$\Rightarrow \boxed{w = \left(X^T X\right)^{-1} (}$$

$$\boxed{W = (X^T X)^{-1} X^T Y}$$

where $X = \begin{bmatrix} \leftarrow x_1^T \rightarrow \\ \vdots \\ \leftarrow x_n^T \rightarrow \end{bmatrix}_{n \times (d+1)}$, $Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$, $W = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_d \end{bmatrix}_{(d+1)}$

$$\boxed{X^T Y)}$$

$)x|$