Linear Regression

$\mathbf{x} \longrightarrow y$

$\mathbf{x}$ is a vector

$\mathbf{x} \in \mathbb{R}^d \rightarrow$ $\mathbf{x}$ is a vector of length d

y is a scalar

$y \in \mathbb{R}$

Prediction or Regression

Predict future income

Current GPA , # AI courses taken	
	$y = 7000$
	$y = 300000$

$\boxed{3.8 \mid 4}$ $\quad y$?

Training data

GPA , # AI	Income
	100000

3.7,	2	100000
3.9,	6	500000
2.4,	1	1,000,000

3.5, 2	?

Functional models.

$$y = f(\mathbf{x})$$

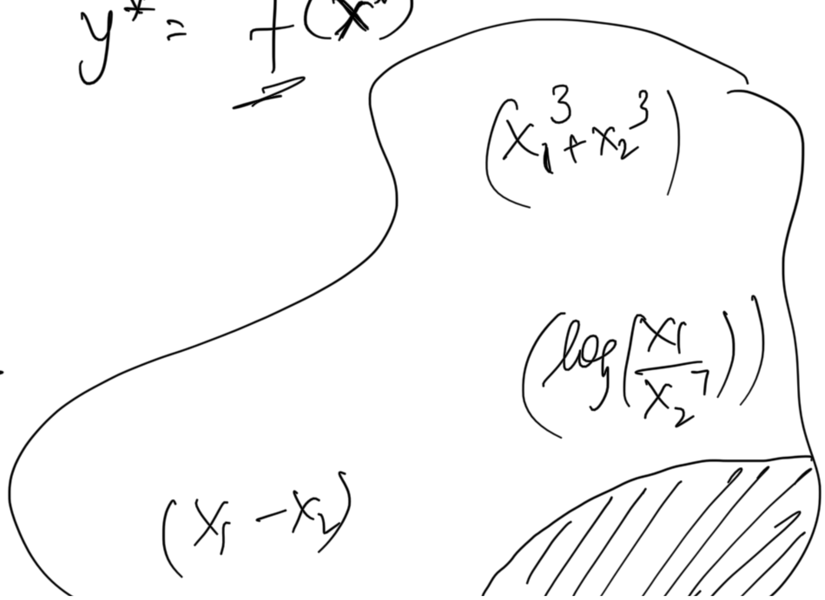
some function.

If we learn $f(\,)$,

then for a new $\mathbf{x}^*$ $\mathbf{x} = \boxed{x_1, x_2}$

$$y^* = \underline{f}(\mathbf{x}^*)$$

a big bag of functions



Inductive bias

Monday Feb 8

$$\mathbf{X} \longrightarrow Y$$

Functional models

$$Y = f(\mathbf{X})$$

Probabilistic Models.

$$p(\mathbf{X}, Y)$$

$p(Y|\mathbf{X}) \leftarrow$ Bayes Rule

$$\mathbf{X} \rightarrow \overset{X_1}{\boxed{x_1 \mid x_2 \mid \; \mid \; \mid \; \mid \; \mid \; \mid}}\ \cdots\ x_D$$

$\mathbf{X}, \mathbf{Y}, \mathbf{Z}$

$\mathbf{X}_1 \quad \mathbf{X}_2$ $\qquad \boxed{X_{11} \quad X_{12} \quad \cdots}$

$$\sum^{D}$$

$$\mathbf{x} \cdot \mathbf{y} = \sum_{i=1} x_i y_i$$

$$= x_1 y_1 + x_2 y_2 + \cdots + x_D y_D$$

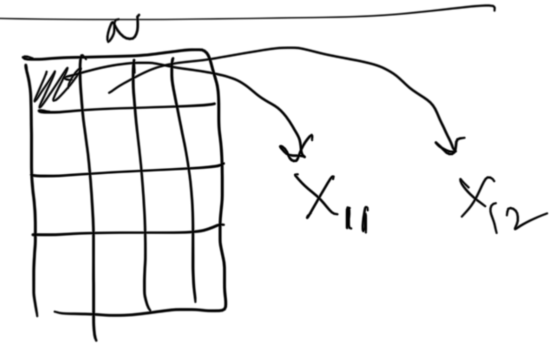
$$|\mathbf{x}| = \sum_{i=1}^{D} |\mathbf{x}_i|$$

$$\|\mathbf{x}\|_2 = \sqrt{\sum_{i=1}^{D} x_i^2} \quad \ell_2 \text{ norm}$$

$$\|\mathbf{x}\|$$

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^{D} x_i^{\,p} \right)^{1/p}$$

<u>Matrix</u>

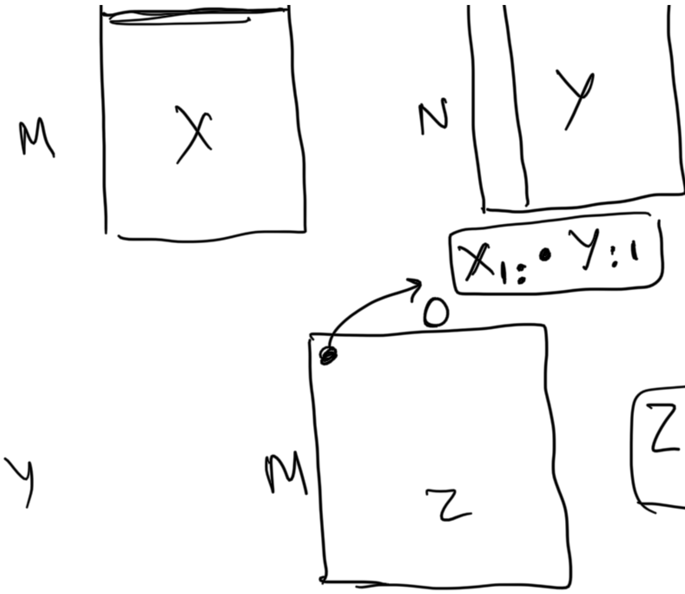


$$\boxed{\mathbf{X} \in \mathbb{R}^{M \times N}}$$

$$Y = X^T \quad (N \times M)$$

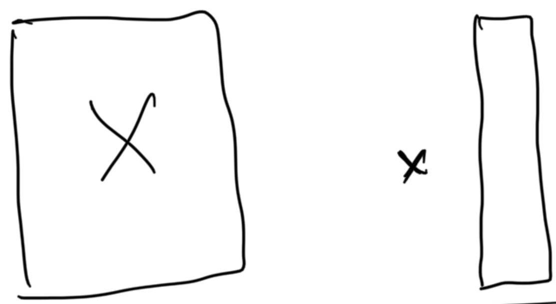
$$c\mathbf{X} = \begin{bmatrix} cx_{11} & cx_{12} & \cdots \\ & \ddots & \\ & & \end{bmatrix}$$

$\mathbf{X}\mathbf{Y}$



$Z = XY$

$$\boxed{Z_{ij} = X_{i:} \cdot Y_{:j}}$$



A D-length vector is a $(D \times 1)$ matrix

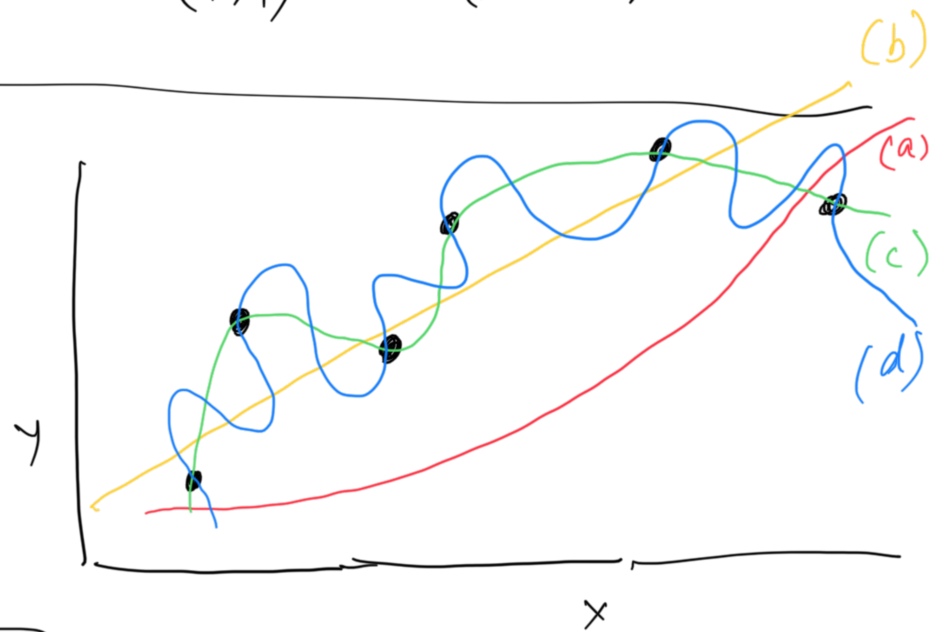
$$x \cdot y = x^T y$$

$1 \times D$

y

$D \times 1$

$$z = x^T y = (1 \times 1) = (x^T) \cdot y$$



Error

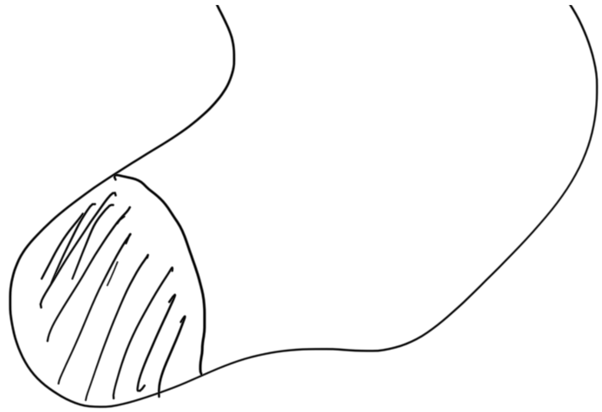
$$(b) \rightarrow \quad y = mx + c$$

$$(c) \rightarrow \quad y = ax^3 + bx^2 + cx + d$$

$$(d) \rightarrow \quad y = ax^{10} + bx^9 + \text{-----} + \text{—}$$

Linear Regression

$y = f(\mathbf{x})$



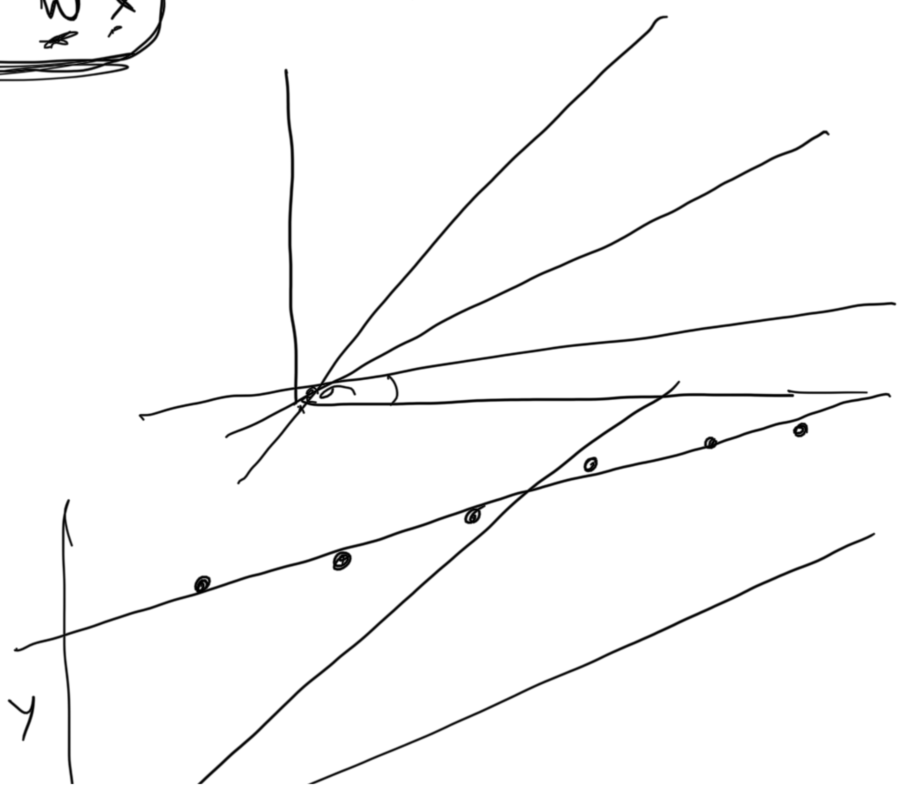
$y = \underline{\mathbf{w}^T \mathbf{x}} \rightarrow D \times 1$

weight vector
$(D \times 1)$

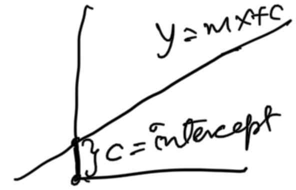
$= w_1 x_1 + w_2 x_2 + \cdots + w_D x_D$

For the flu example

$y = w x$



y=w



$$\boxed{y = w_0 + w_1 x}$$

bias-term

Given some data:

x	y
12	9
28	15
15	11
48	21
56	22

Find the best w_0, w_1 which "fits" the data best.