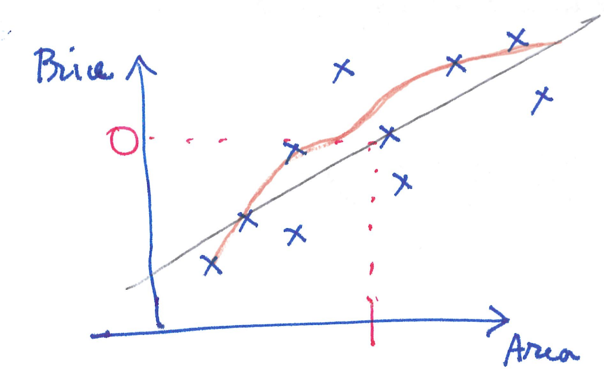
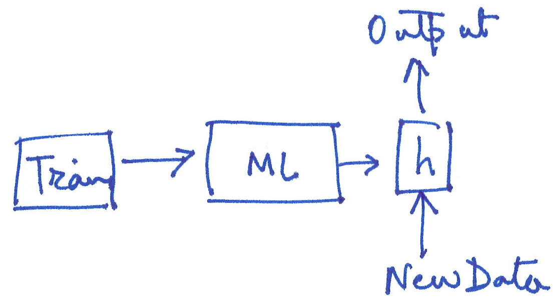
CS 307 Lecture 2

Housing market.

Area	# Bed	Price ($1000)
2104	3	400
1600	3	330
2400	3	369
1416	2	232
3000	4	540
⋮		⋮





Supervised Learning

Training Data

- Input / features $\in \mathbb{R}^d \} \mathcal{X}$
- Output $\in \mathbb{R}$ (price of the house) $\} \mathcal{Y}$

Goal: Predict output on new data.

Hypothesis: $h : \mathcal{X} \to \mathcal{Y}$

- Regression: $\mathcal{Y} = \mathbb{R}$
- Classification: $\mathcal{Y}$ = finite set.

Unsupervised

Training Data

- Input / features

Linear Regression:

Hypothesis is a linear function of the features.

Training data = $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(1)}), \ldots (x^{(n)}, y^{(n)})\}$

Input features $\in \mathbb{R}^d$

Output $\in \mathbb{R}$

$h : \mathbb{R}^d \to \mathbb{R}$ is of the form

$$h_\theta(x) = \theta_0 + \theta_1 x_1 + \cdots + \theta_d x_d$$

$x = (x_1, x_2 \ldots x_d)$

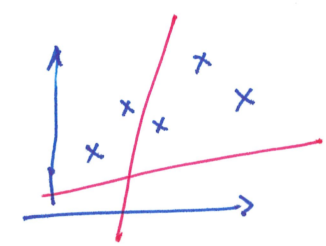
Example: Housing market.

Training Data = $\{((2104, 3), 400), ((1600, 3), 330) \ldots \}$

$x^{(1)} = (x_1^{(1)}, x_2^{(1)})$

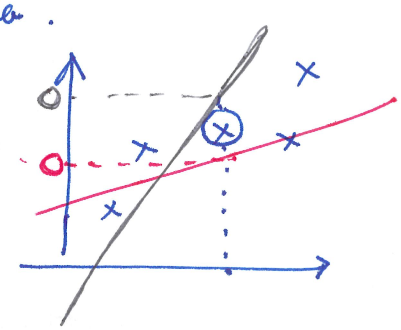
$y^{(1)} =$

$$h_\theta = \theta_0 + \theta_1 x_1 + \theta_2 x_2. \quad | \; d = 2.$$



Cost function: Characterizes how well a hypothesis performs on the training data.

$$J(\theta) = \frac{1}{2} \sum_{i=1}^{n} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$



Problem: Find $\theta \in \mathbb{R}^{d+1}$ that minimizes

$$J(\theta) = \frac{1}{2} \sum_{i=1}^{n} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

Matrices: $A \in \mathbb{R}^{n \times d}$ A_{ij} — (i,j) the entry of A.

$A^T \in \mathbb{R}^{d \times n}$: $(A^T)_{ij} = A_{ji}$

$$A = \begin{bmatrix} A_{11} & A_{12} & \cdots \\ A_{21} & \cdot & \\ \vdots & & \ddots \end{bmatrix} \quad A^T = \begin{bmatrix} A_{11} & A_{21} & \cdots \\ A_{12} & & \\ \vdots & & \end{bmatrix}$$

Vectors will be thought of as column vectors

$x \in \mathbb{R}^d \longrightarrow$ Think of as $x \in \mathbb{R}^{d \times 1}$

$$x = x_1 \ldots x_d \longrightarrow \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_d \end{bmatrix}$$

For $x \in \mathbb{R}^d$

$$\|x\|_2 = \sqrt{\sum_{i=1}^{d} x_i^2} \quad ; \quad \|x\|_2^2 = \sum_{i=1}^{d} x_i^2 .$$

Question: $x \in \mathbb{R}^{d \times 1}$. Which of the following is well defined?

(a) $x x$ ✗
(b) $x^T x$ — $\in \mathbb{R}$
(c) $x x^T$ — $d \times d$
(d) $x^T x^T$ ✗

$A \in \mathbb{R}^{n \times d}$
$B \in \mathbb{R}^{c \times m}$
$A \times B$ — defined $c = d$
⟶ when defined $n \times m$.

$x^T = 1 \times d$ $x = d \times 1$
$x^T x = 1 \times 1$, $x x^T = d \times d$

$$\|x\|_2^2 = \sum_{i=1}^{d} x_i^2 = x^T x .$$

Partial derivative: $f : \mathbb{R}^d \longrightarrow \mathbb{R}$
$\longrightarrow (\theta_1, \theta_2 \ldots \theta_d)$

$\dfrac{\partial f}{\partial \theta_i}$ — How f changes while θ_i changes (while $\theta_1 \ldots \theta_{i-1}, \theta_{i+1}, \ldots \theta_d$ are fixed.)

Example. $f(\theta) = \theta_1 + 2\theta_2 + 3\theta_3^2$

$$\frac{\partial f}{\partial \theta_1} = 1 \qquad \frac{\partial f}{\partial \theta_3} = 3(2\theta_3) = 6\theta_3 .$$

Gradient : $f: \mathbb{R}^d \to \mathbb{R}$. (arguments θ).

$$\frac{\partial f}{\partial \theta_1}, \frac{\partial f}{\partial \theta_2} \cdots \frac{\partial f}{\partial \theta_d}.$$

$$\nabla_\theta f = \begin{bmatrix} \frac{\partial f}{\partial \theta_1} \\ \frac{\partial f}{\partial \theta_2} \\ \vdots \\ \frac{\partial f}{\partial \theta_d} \end{bmatrix}$$

Example : $f: \theta_1 + 2\theta_2 + 3\theta_3^2$

$$\nabla_\theta f = \begin{bmatrix} 1 \\ 2 \\ 6\theta_3 \end{bmatrix}$$

Example : $f(\theta) = \theta_1 x_1 + \theta_2 x_2 + \cdots \theta_d x_d$.

$$= x^T \theta = \theta^T x.$$

$$\frac{\partial f}{\partial \theta_i} = x_i$$

$$\nabla_\theta f = \begin{bmatrix} \frac{\partial f}{\partial \theta_1} \\ \vdots \\ \frac{\partial f}{\partial \theta_d} \end{bmatrix} = x \qquad \nabla_\theta \, x^T \theta = x.$$

$$f(\theta) = \theta^T A \theta \quad \text{where} \quad \theta \in \mathbb{R}^{d \times 1}, \; A \in \mathbb{R}^{d \times d}$$

#row = d

#columns = d.

$$\underbrace{\nabla_\theta \, \theta^T A \theta}_{\underbrace{a\theta^2}} = (A + A^T)\,\theta.$$

$$a\theta^2 \to 2a\theta.$$