Plan

Review

Logistic Regression

↳ Sigmoid

↳ Softmax

Cross Entropy

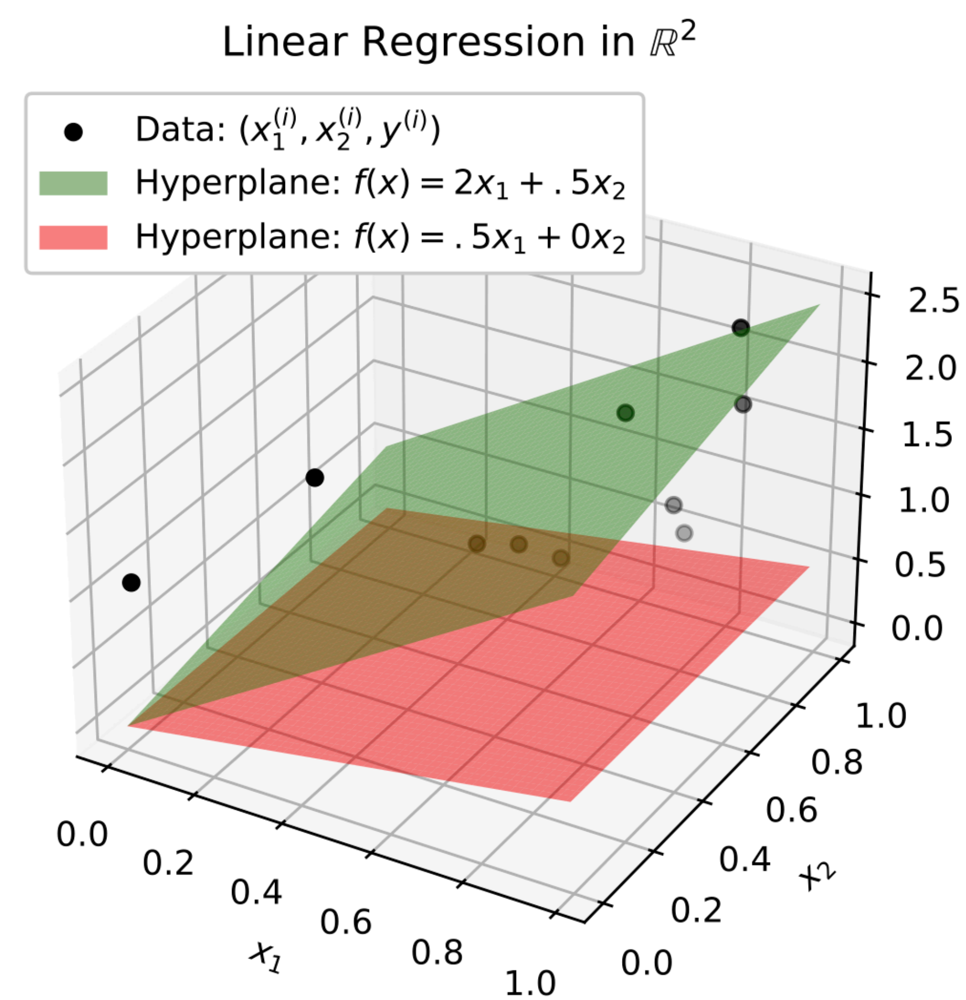
Logistics

- Check in
- Scribed notes
- 2-3 work and struggle

Linear Regression

$(x^{(1)}, y^{(1)}), \ldots, (x^{(n)}, y^{(n)})$

$x^{(i)} \in \mathbb{R}^d \qquad y^{(i)} \in \mathbb{R}$



① Model: $f(x) = w \cdot x$

for $w \in \mathbb{R}^d$

$X \in \mathbb{R}^{n \times d}$

$y \in \mathbb{R}^n$

② Loss: $\mathcal{L}(w) = \frac{1}{n} \| Xw - y \|_2^2$

$$= \frac{1}{n} \sum_{i=1}^{n} (\underset{1 \times d}{w^T} \underset{d \times 1}{x^{(i)}} - \underset{1}{y^{(i)}})^2$$

$\nabla_w \mathcal{L}(w^*) = 0$

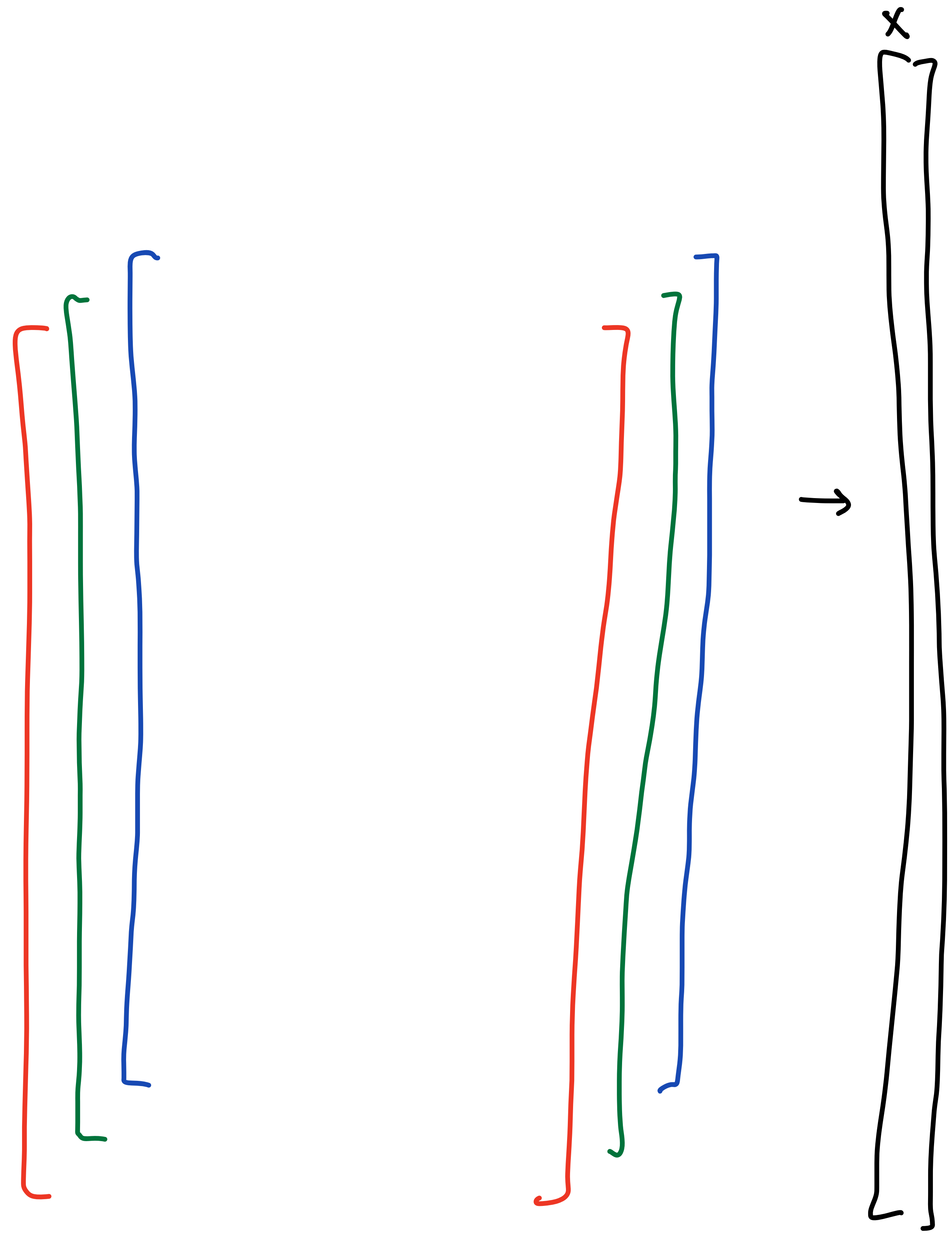
$\Longleftrightarrow w^* = (\underset{d \times n}{X^T} \underset{n \times d}{X})^{-1} \underset{d \times n}{X^T} \underset{n \times 1}{y}$

$d \times d \qquad d \times 1$

$d \times 1$

③ Optimizer: $\nabla_w \mathcal{L}(w) = \begin{bmatrix} \vdots \\ \frac{\partial \mathcal{L}}{\partial w_j} \\ \vdots \end{bmatrix}$

Motivation



$\cancel{y \in \mathbb{R}}$

$y \in \{0,1\}$

Is there a cat in this image?

(x, y)

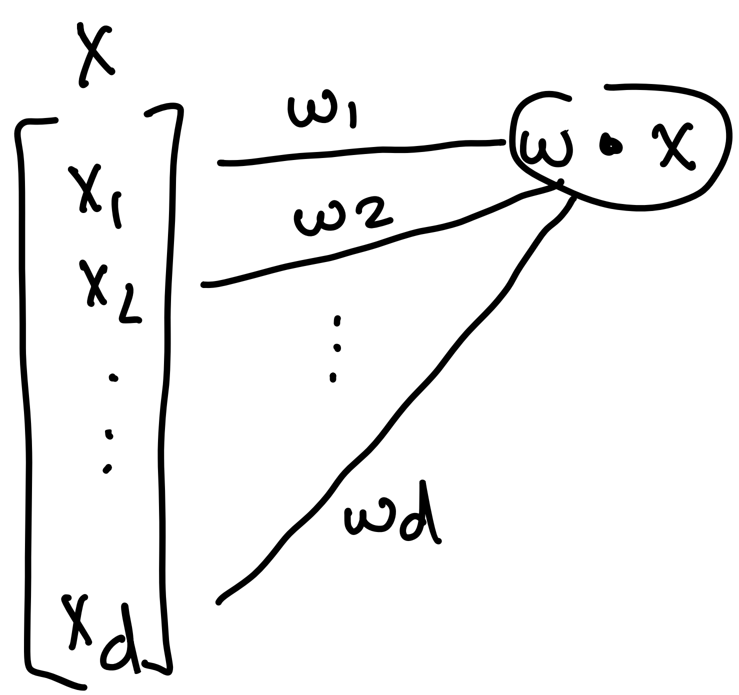
Supervised Binary Classification

$\ldots (x^{(i)}, y^{(i)}) \ldots$

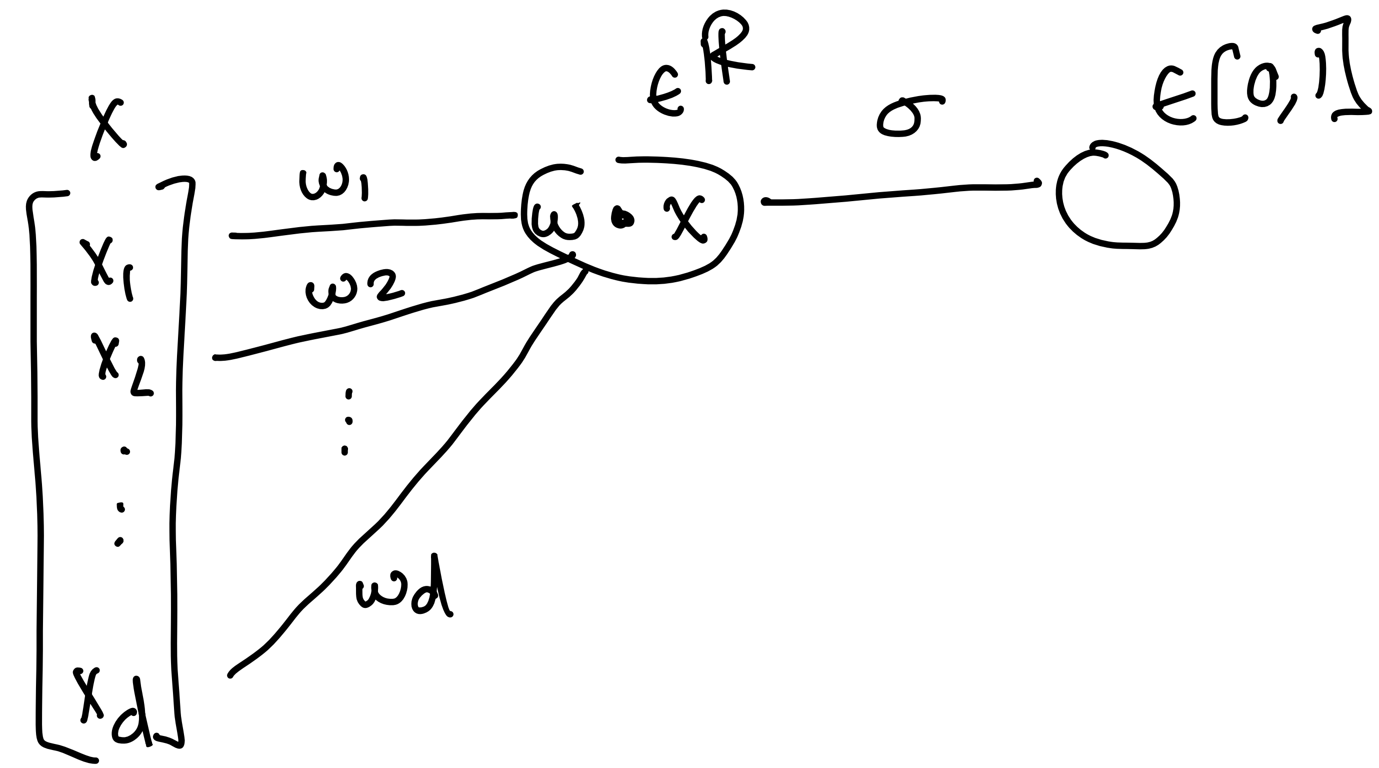
$x^{(i)} \in \mathbb{R}^d \qquad y^{(i)} \in \{0,1\}$

$w \in \mathbb{R}^d$

Goal: $f(x)$ = probability of positive class



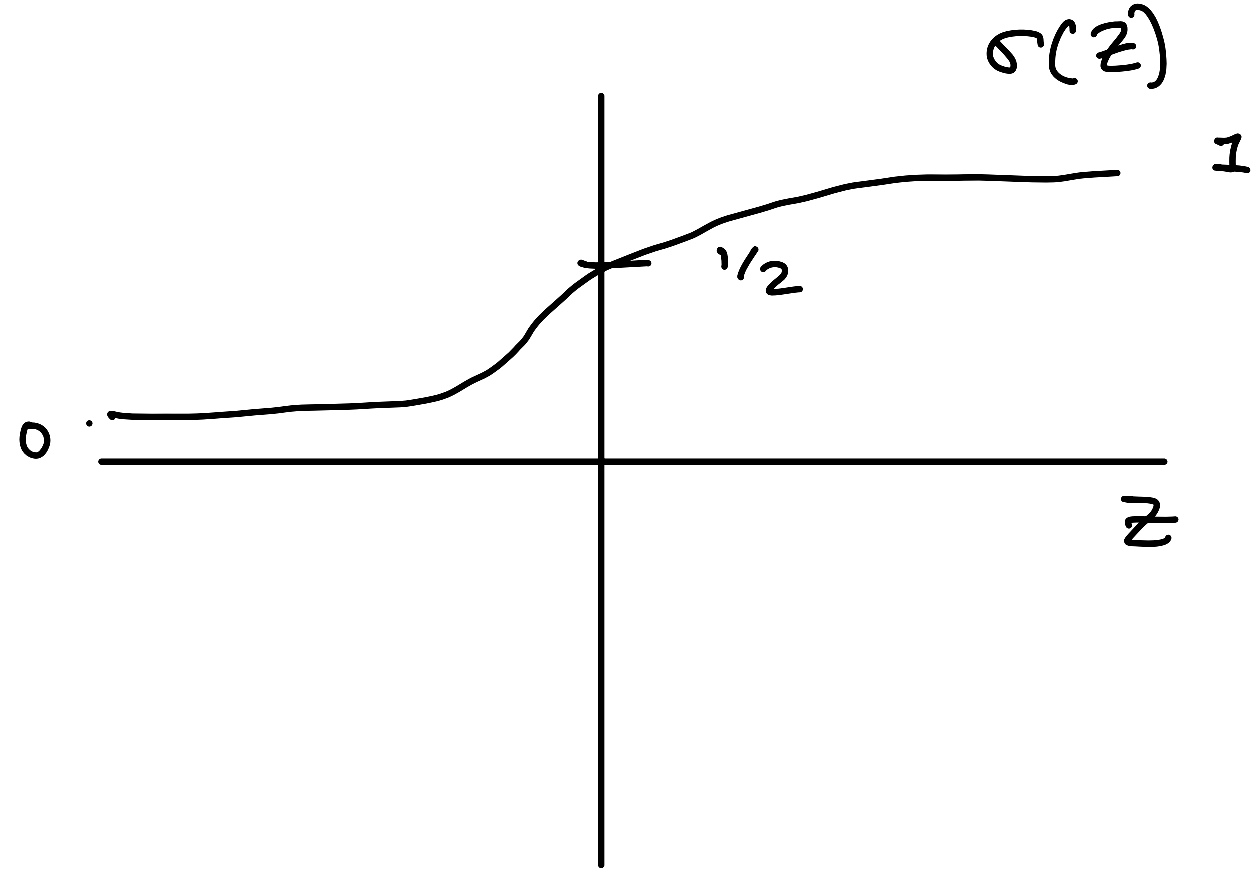
Linear Regression

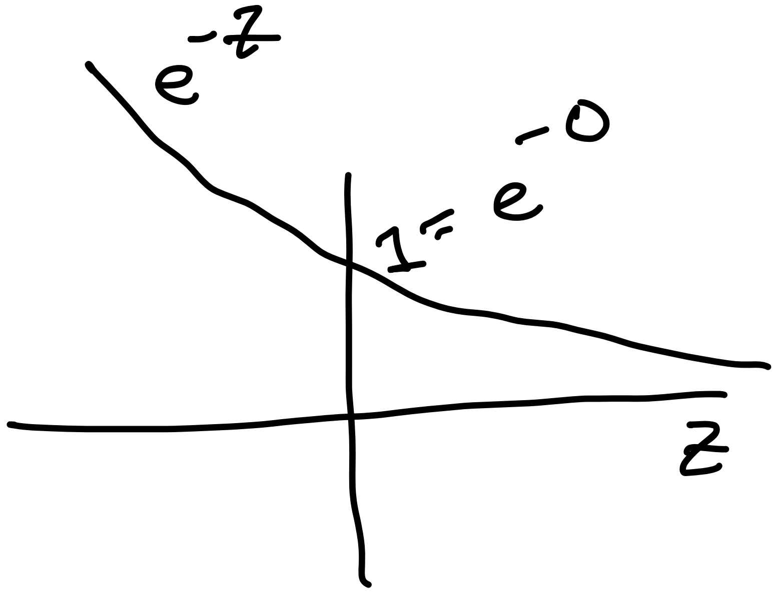


Logistic Regression

Sigmoid

$$\sigma(z) = \frac{1}{1+e^{-z}}$$



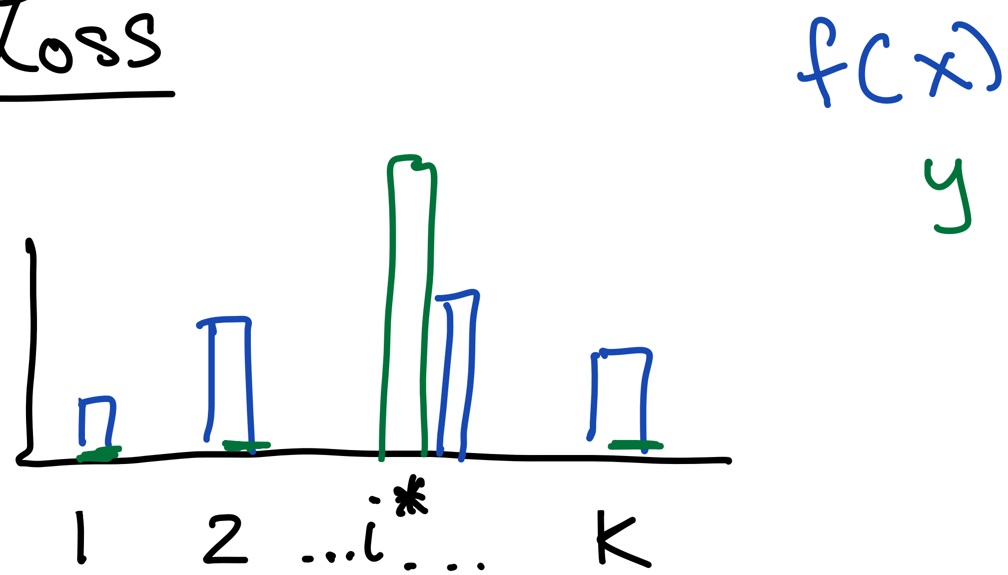


$$\lim_{z \to \infty} \frac{1}{1+e^{-z}} = \frac{1}{1+0} = 1$$

$$\lim_{z \to -\infty} \frac{1}{1+e^{-z}} = \frac{1}{\infty} = 0$$

Loss

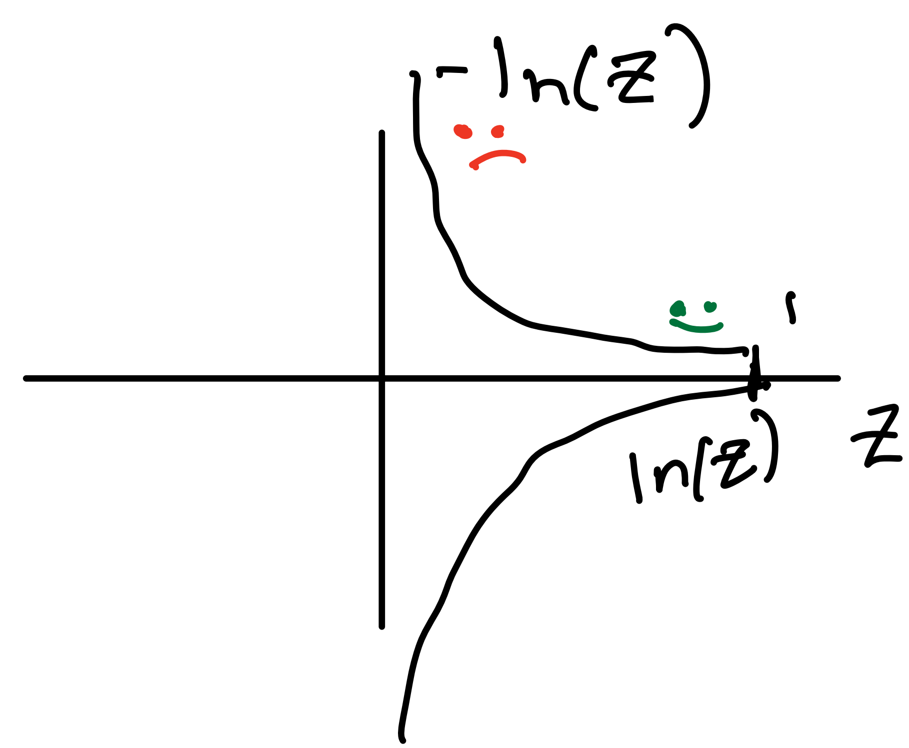
$f: \mathbb{R}^d \to [0,1]^K$



Goal: Measure distance between distribution $f(x)$ and y

$$H(y, f(x)) = -\sum_{i=1}^{K} y_i \log_2 f(x)_i = -\log_2 f(x)_{i^*}$$

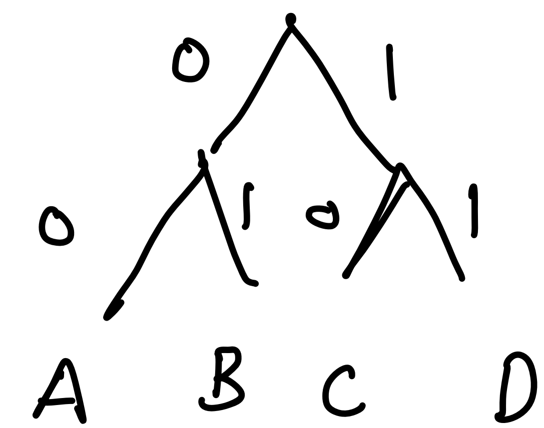
$$y_i = \begin{cases} 1 & \text{if } i = i^* \\ 0 & \text{else} \end{cases}$$



Cross Entropy

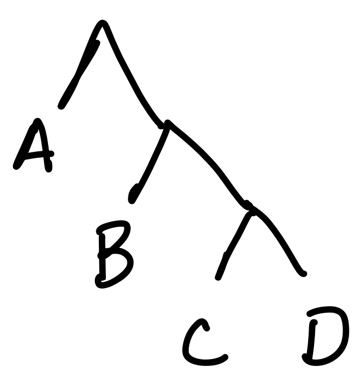
Communicate A, B, C, D

Approach #1: 00, 01, 10, 11



What if I know distribution?

e.g. $q_A = 1/2$, $q_B = 1/4$, $q_C = 1/8$, $q_D = 1/8$

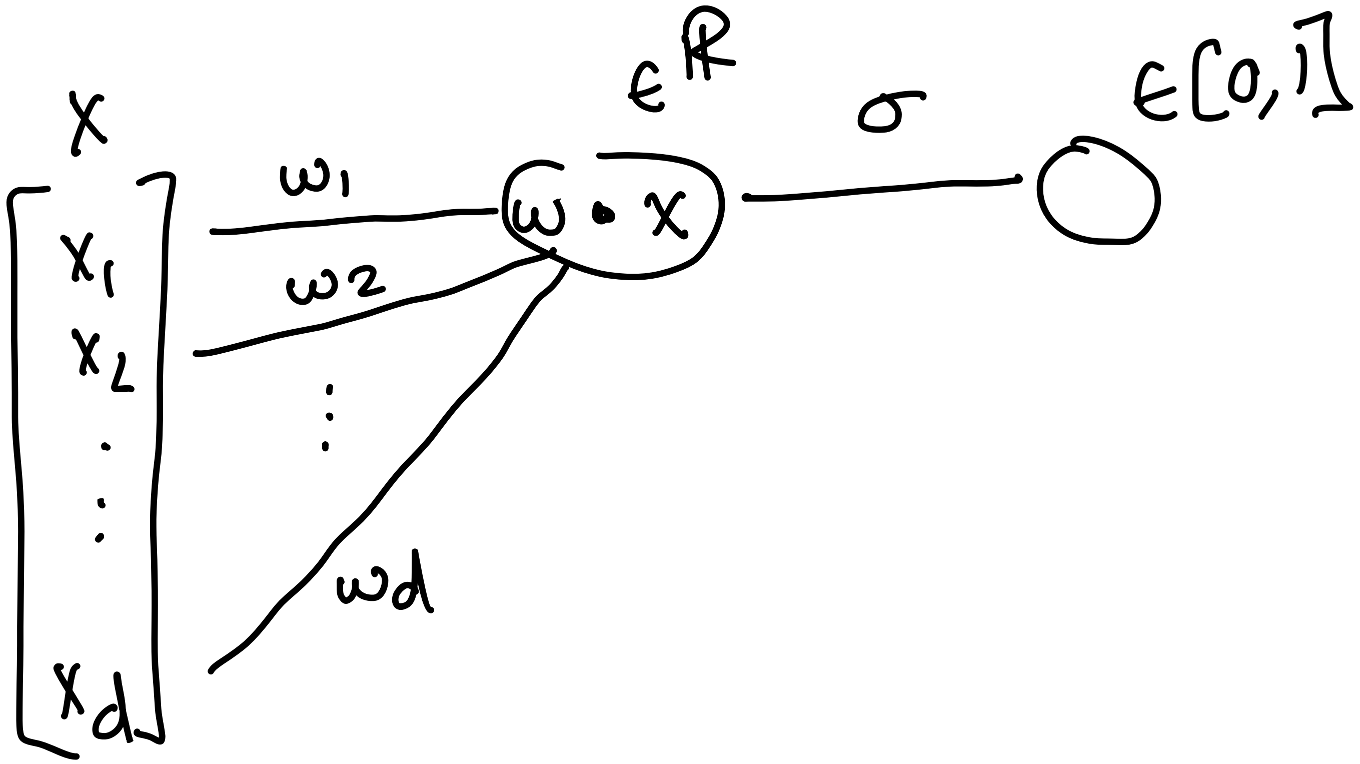


$$l_i = -\log_2 q_i$$
$$= \log_2 (1/q_i)$$

$$H(q) = \mathbb{E}[\#\,\text{bits}] = \mathbb{E}_{i \sim q}[l_i] = -\mathbb{E}_{i \sim q}[\log_2 q_i] = -\sum_{i=1}^{K} q_i \log_2 (q_i)$$

$H(p, q)$ = communicate w/ diff. dist. $= -\mathbb{E}_{i \sim p}[\log_2 q_i]$

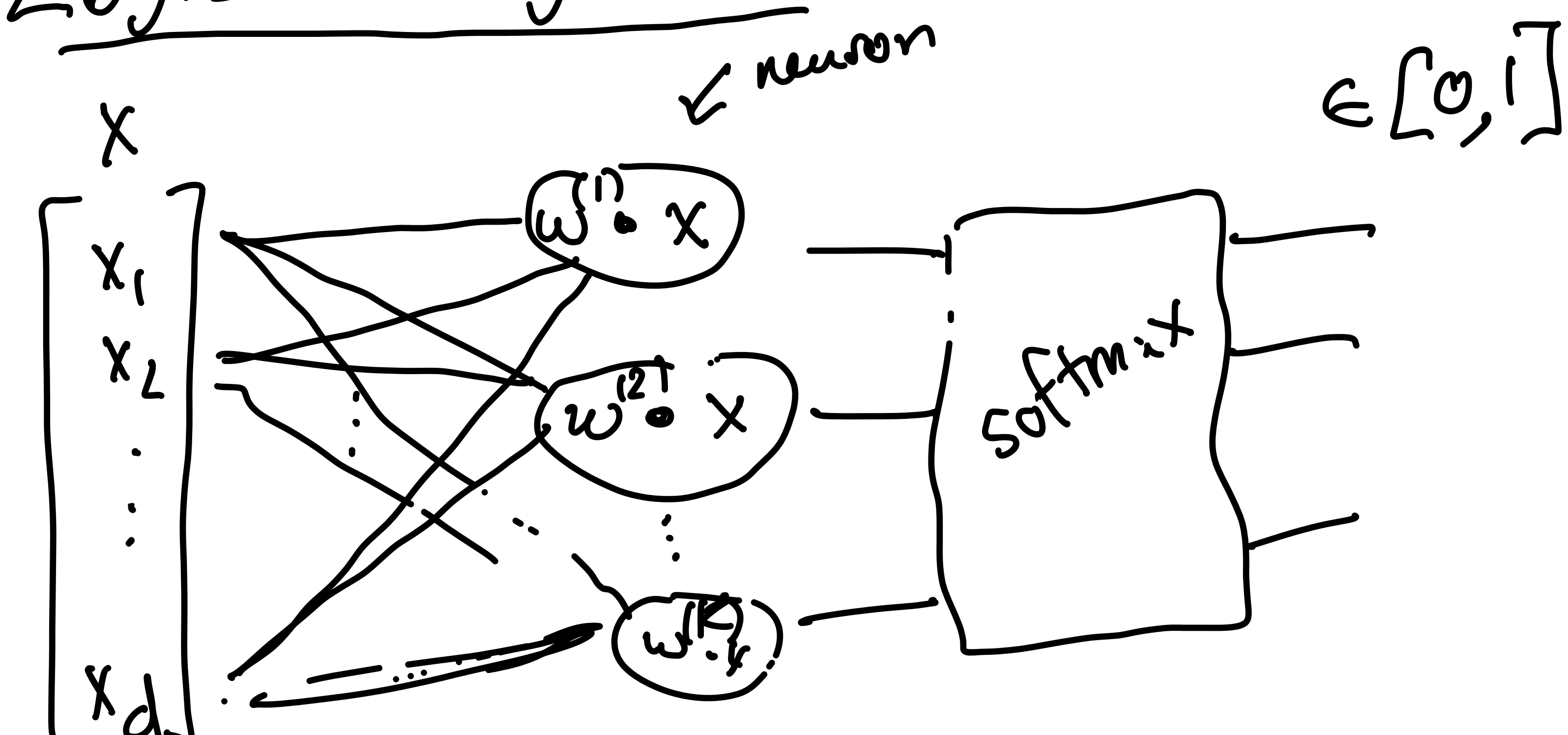
Binary Logistic Regression



$$z \in \mathbb{R}^K$$

$$\text{softmax}(z) =_{norm} \left(\begin{bmatrix} e^{z_1} \\ \vdots \\ e^{z_j} \\ \\ e^{z_K} \end{bmatrix} \right) = \frac{1}{\sum_{\ell=1}^{K} e^{z_\ell}} [\ ]$$

Logistic Regression



$$\sum_{j=1}^{K} \text{softmax}(z)_j$$

$$= \sum_{j=1}^{K} e^{z_j} \Big/ \sum_{\ell=1}^{K} e^{z_\ell}$$

$$= \frac{\sum_{j=1}^{K} e^{z_j}}{\sum_{\ell=1}^{K} e^{z_\ell}} = 1$$