Log. Reg.

Given x, want $\hat{y} = P(y=1|x)$, $0 \le \hat{y} \le 1$

$x \in \mathbb{R}^{n_x}$

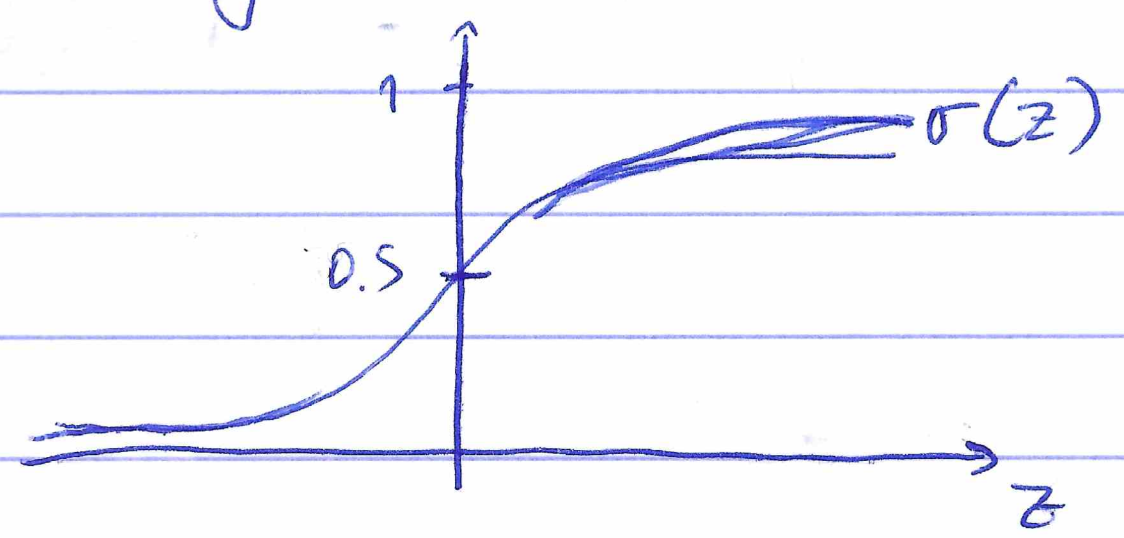
Parameters: $w \in \mathbb{R}^{n_x}$, $b \in \mathbb{R}$

Output: ~~$\hat{y} = w^T x + b$~~ (lin. reg., not good for bin. class.)

so: $\hat{y}^{(i)} = \sigma(w^T x^{(i)} + b)$

other notation, not used here

$\boxed{\text{sometimes: } x_0 = 1,\ x \in \mathbb{R}^{n_x+1},\quad \hat{y} = \sigma(\theta^T x),\quad \theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{n_x} \end{bmatrix} \begin{matrix} \}\,b \\ \\ \Big\}\,w \\ \\ \end{matrix}}$



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

If z large, $\sigma(z) \approx \frac{1}{1+0} = 1$

" very -ve, $\sigma(z) \approx \frac{1}{1+\text{bignum}} \approx 0$

Given $\{(x^{(1)}, y^{(1)}), \ldots, (x^{(m)}, y^{(m)})\}$, want $\hat{y}^{(i)} \approx y^{(i)}$

$z^{(i)} = w^T x^{(i)} + b$

Loss (error) function: ~~$L(\hat{y}, y) = \frac{1}{2}(\hat{y} - y)^2$~~

applied to a single example

$$L(\hat{y}, y) = -(y \log \hat{y} + (1-y) \log(1-\hat{y}))$$

If $y=1$: $L(\hat{y}, y) = -\log \hat{y}$; want $\log \hat{y}$ large, want $\hat{y}$ large

" $y=0$: $L(\hat{y}, y) = -\log(1-\hat{y})$; " $\log(1-\hat{y})$ ", want $\hat{y}$ small

applied to all (avg of LFs of entire train. set)

C.F.: $J(w,b) = \frac{1}{m} \sum_{i=1}^{m} L(\hat{y}^{(i)}, y^{(i)})$

$$= -\frac{1}{m} \sum_{i=1}^{m} \left[y^{(i)} \log \hat{y}^{(i)} + (1-y^{(i)}) \log(1-\hat{y}^{(i)}) \right]$$