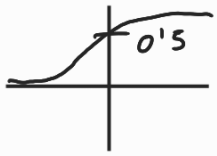
- Given x, want $\hat{y} = P(y=1 \mid x)$; $x \in \mathbb{R}^{n_x}$
- Parameters: $w \in \mathbb{R}^{n_x}$, $b \in \mathbb{R}$
- Output : $\boxed{\hat{y}^{(i)} = \sigma(w^T x^{(i)} + b)}$ $\Rightarrow 0 \leq \hat{y} \leq 1$



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

- Error function : $-(y \log \hat{y} + (1-y) \log (1-\hat{y}))$
  - If $y=1$; $L(\hat{y}, y) = -\log \hat{y}$ (we want $\hat{y}$ to be as big as possible)
  - If $y=0$; $L(\hat{y}, y) = -y \log (\hat{y})$ (we want $\hat{y}$ to be as small as possible)
- Cost function : $J(w,b) = \min\left(\frac{1}{m} \sum_{i=1}^{m} L(\hat{y}^{(i)}, y)\right)$

- GRADIENT DESCENT :