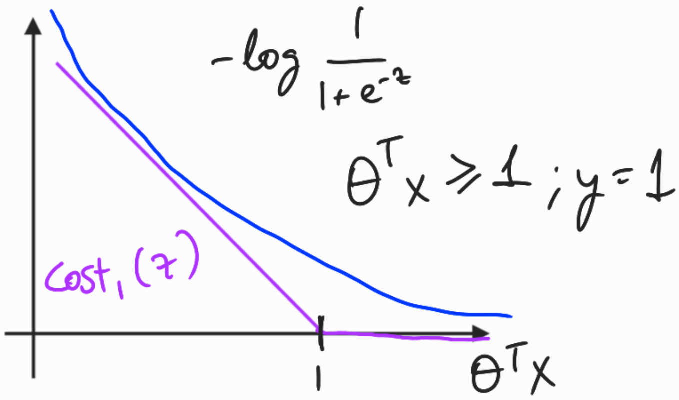
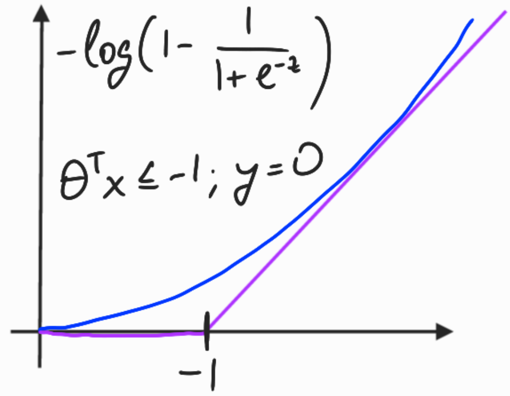




$$J(\theta) = \min_{\theta} \cancel{\frac{1}{m}} \sum_{i=1}^{m} y^{(i)} cost_1(\theta^T x^{(i)}) + (1-y^{(i)}) cost_0(\theta^T x^{(i)}) + \frac{\lambda}{2\cancel{m}} \sum_{i=0}^{n} \theta_j^2$$

$$= \min_{\theta} \underset{\substack{\downarrow \\ 1/\lambda}}{C} \sum_{i=1}^{m} \left[y^{(i)} cost_1(\theta^T x^{(i)}) + (1-y^{(i)}) cost_0(\theta^T x^{(i)}) \right] + \frac{1}{2} \sum_{i=1}^{n} \theta_j^2$$

$$h_\theta(x) \begin{cases} 1 & \text{if } \theta^T x \geq 0 \\ 0 & \text{if } \theta^T x < 0 \end{cases}$$

$$J(\theta) = C \cdot A + B$$

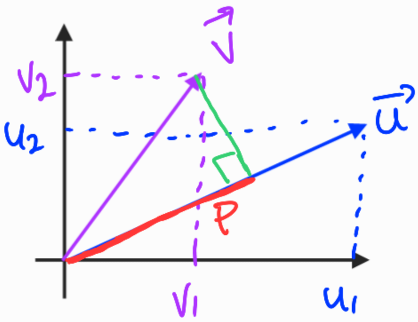
C ↗ large ⇒ no missclassification allowed easily
↓ bias, ↑ variance

C ↘ small ⇒ missclassification allowed easily
↑ bias, ↓ variance

LARGE MARGIN CLASSIFIERS

$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ $v = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$

Reminder: Vector inner product $\rightarrow u^T v$



$$u^T v = p \cdot \|u\| = u_1 v_1 + u_2 v_2 \;;\; p \in \mathbb{R}$$