2.18 logistic cost function

If $y=1$: $P(y|x) = \hat{y}$

If $y=0$: $P(y|x) = 1 - \hat{y}$

$$P(y|x) = \hat{y}^{y} (1-\hat{y})^{(1-y)}$$

$$\max \left(\log(P(y|x)) = \log \left(\hat{y}^{y} (\hat{y})^{(1-y)} \right) \right)$$

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

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

$$\min(-\log(P(y|x))) = - \left(y \log \hat{y} + (1-y) \log(1-\hat{y}) \right)$$

$$= L(\hat{y}, y)$$

3.6 激活函数的选择

sigmoid function : $a = \dfrac{1}{1+e^{-z}}$

$a = \tanh(z) = \dfrac{e^{z} - e^{-z}}{e^{z} + e^{-z}}$ (better)

ReLU : $a = \max(0, z)$

2.11 向量化 vectorization.

$$X = \begin{bmatrix} x^{(1)} & x^{(2)} & \cdots & x^{(m)} \\ | & | & & | \end{bmatrix} \quad Y = [y^{(1)}, y^{(2)}, \cdots y^{(m)}]$$

$$Z = [Z^{(1)} \; Z^{(2)} \cdots Z^{(m)}] = W^{T} X + [b \; b \sim b]$$

$$Z = np.dot(W^{T}, X) + b$$

$$A = [a^{(1)} \cdots a^{(m)}] \quad Y = [y^{(1)} \cdots y^{(m)}]$$

$$dz = A - Y$$

2.15 Broadcasting

use np.random.randn(1,5)

but not np.random.randn(5)

Can use reshape to set the dim to
the matrix.