#2.3 Logistic Regression Cost Function

↑ can be viewed as a small NN.!

$$\hat{y}^{(i)} = \sigma(w^T x^{(i)} + b), \text{ where } \sigma(z^{(i)}) = \frac{1}{1+e^{-z^{(i)}}}$$

Q: Given $\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \dots, (x^{(m)}, y^{(m)})\}$,

want $\hat{y}^{(i)} \approx y^{(i)}$.

$\hat{y}^{(i)}$: predict; $y^{(i)}$: real

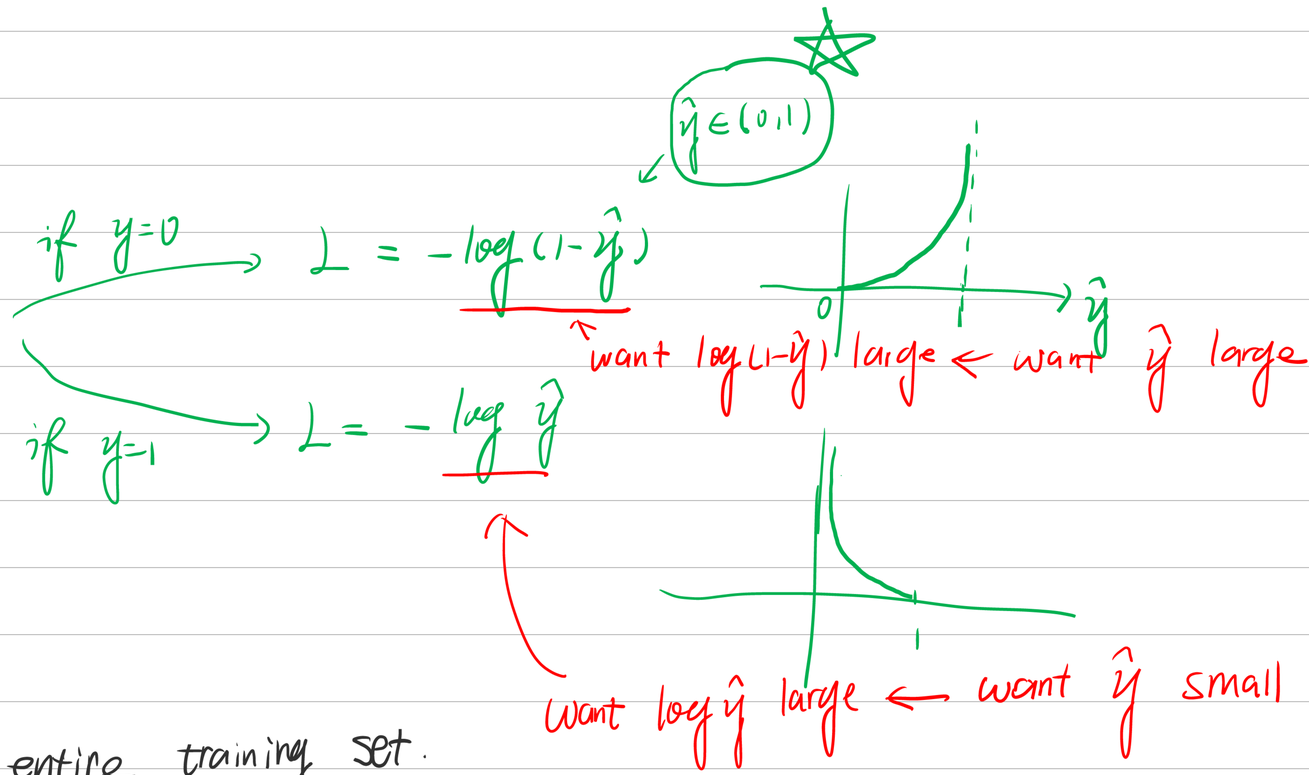
Loss (error) function:

↑ want the loss ↓ as small as possible.

MSE: $L(\hat{y}, y) = \frac{1}{2}(\hat{y} - y)^2$

↑ error

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



Cost function

↑ measure the performance on an entire training set.

- MSE: $J(w,b) = \frac{1}{m}\sum_{i=1}^{m} L(\hat{y}^{(i)}, y^{(i)}) = \frac{1}{m}\sum_{i=1}^{m}(\hat{y}^{(i)} - y^{(i)})^2$

↑ error

- $J(w,b) = \frac{1}{m}\sum_{i=1}^{m} L(\hat{y}^{(i)}, y^{(i)}) = -\frac{1}{m}\sum_{i=1}^{m}[\, y \log \hat{y} + (1-y)\log(1-\hat{y}) \,]$