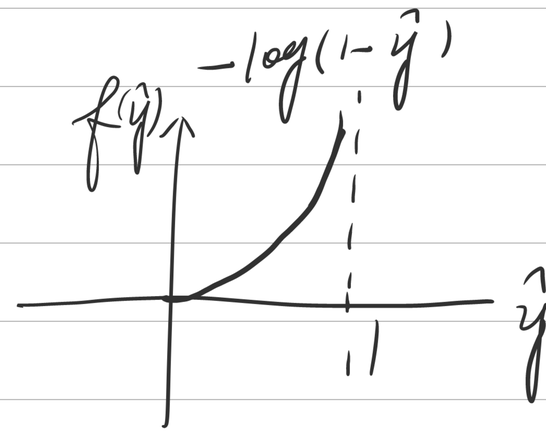
2.4 Gradient descent

Recap:

- Predicted value: $\hat{y} = \sigma(w^T x + b)$

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

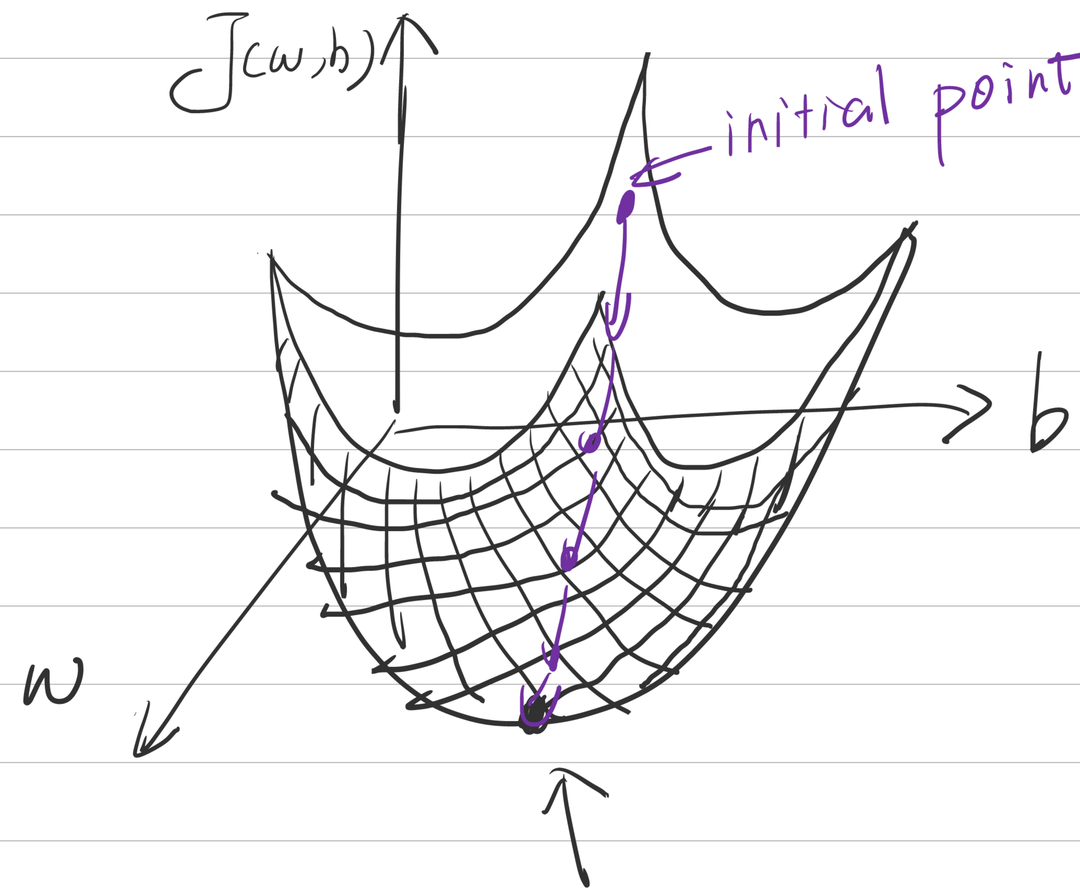
- Loss function: $L(\hat{y}, y) = -[y \log \hat{y} + (1-y)\log(1-\hat{y})]$



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

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

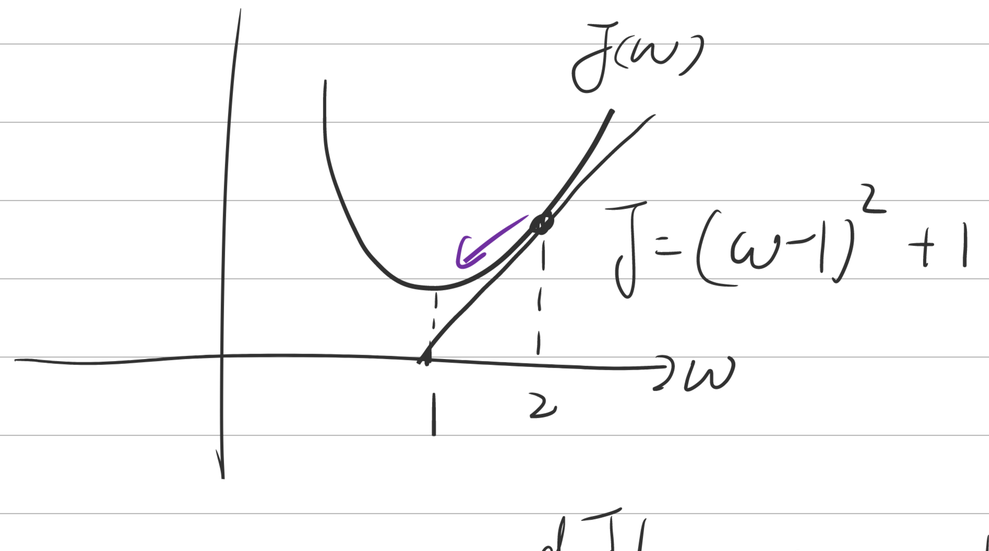
- Goal: Want to find w, b that minimize $J(w,b)$.



- Algorithm:

partial derivative / d: derivative

$$\begin{cases} w := w - \alpha \frac{\partial J}{\partial w} \\ b := b - \alpha \frac{\partial J}{\partial b} \end{cases}$$



$$\left.\frac{dJ}{dw}\right|_{w=2} = \left.2(w-1)\right|_{w=2} = 2$$

$$w' = w\Big|_{w=2} - \alpha \cdot \left.\frac{dJ}{dw}\right|_{w=2}$$

$$= 2 - 0.5 \cdot 2$$

$$= 1$$