Sep 24, 2024

$$\begin{array}{lll} x_1 & y_1 & \hat{y}_1 \\ \vdots & \vdots & \vdots \\ x_N & y_N & \hat{y}_N \end{array}$$

ground truth    prediction

goal: $\hat{y}_i$ the same as $y_i$

$$\hat{y}_i = f(x_i)$$

linear regression    $\hat{y}_i = w^T x_i$

logistic regression    $\hat{y}_i = \sigma(w^T x_i)$

loss function

linear regression $J(w) = \sum_{i=1}^{N} (y_i - \hat{y}_i)^2$ squared loss

logistic regression

$$J(w) = \sum_{i=1}^{N} - y_i \log \hat{y}_i - (1-y_i) \log (1-\hat{y}_i)$$

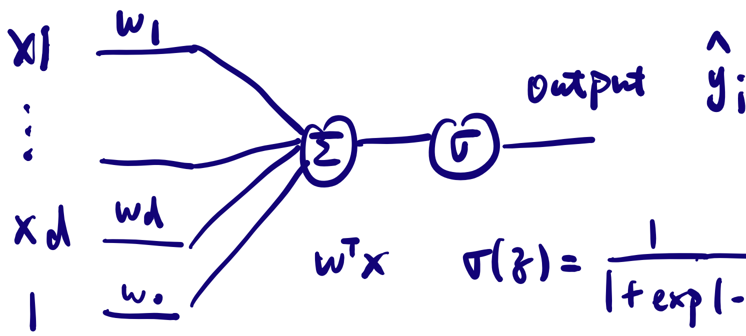
cross entropy / log loss

$$J(w) = \sum_{i=1}^{N} | y_i - \hat{y}_i | + \frac{1}{2} \lambda w^T w$$

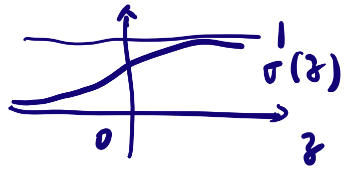
Gradient Descent

$\min J(w)$

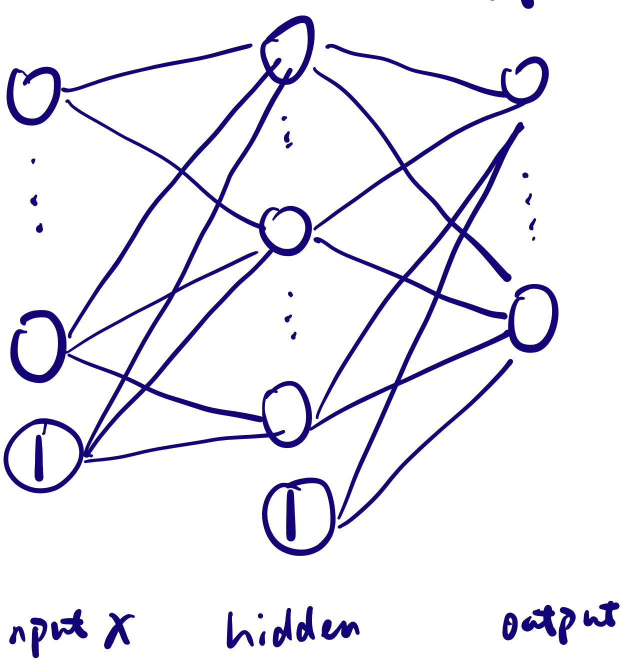
$$w_k = w_{k-1} - \eta \frac{\partial J(w)}{\partial w}$$




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




$$\frac{\partial \sigma(z)}{\partial z} = \sigma(1-\sigma)$$



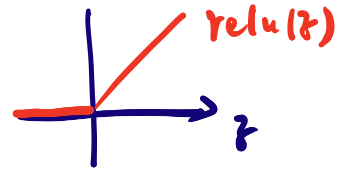

$h = w_h^T x \qquad O = w_o^T h = \underline{w_o^T (w_h^T x)}$ linear function

$h = \sigma(w_h^T x) \qquad O = \sigma(w_o^T h)$

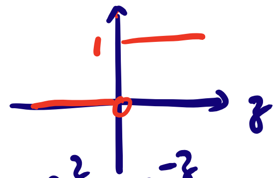
non-linear activation function

sigmoid $\sigma(z)$

$$\mathrm{relu}(z) = \max(0, z)$$




$$\frac{\partial\, \mathrm{relu}(z)}{\partial z}$$




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