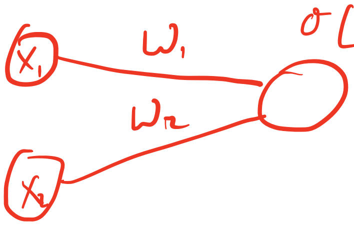
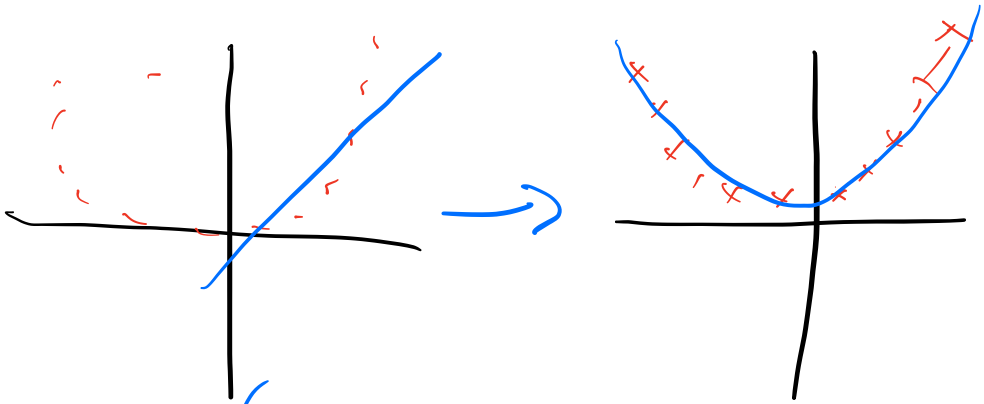


$\sigma(x^T w) = \sigma(x_1 w_1 + x_2 w_2) = P(y=1 \mid x, w)$

Logistic

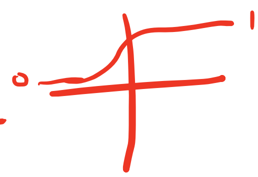


$X = \begin{bmatrix} x_1 \\ x_1 \end{bmatrix}$

$\phi(x) = \begin{bmatrix} x_1 \\ x_1^2 \\ 1 \end{bmatrix} \longrightarrow$ Linear $\phi(x)^T w = x_1 w_1 + x_1^2 w_2 + 1(b)$

Logistic

$P(y=1 \mid x, w) = \sigma(\phi^T w)$



Neural network

$$\phi(x) = \begin{bmatrix} g_1(x) \\ g_2(x) \\ g_3(x) \end{bmatrix} \longrightarrow \begin{bmatrix} \sigma(x^T w_1) \\ \sigma(x^T w_2) \\ \sigma(x^T w_3) \end{bmatrix}$$

neural network regression

$$y = \phi(x)^T w_0 \longrightarrow \underbrace{\sigma(x^T w_1)}_{\phi_1(x)} w_{01} + \underbrace{\sigma(x^T w_2)}_{\phi_2(x)} w_{02} + \underbrace{\sigma(x^T w_3)}_{\phi_3(x)} w_{03}$$

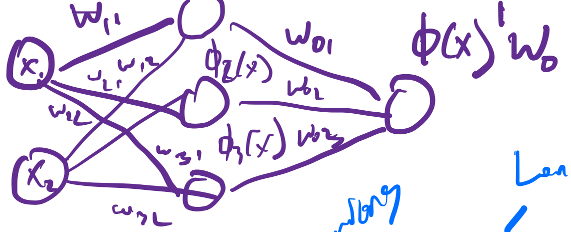
$$= \sigma(x_1 w_{11} + x_2 w_{12} \dots)$$

Parameters

$$X = \begin{array}{c|c} & \text{measurements } (d) \\ \hline \text{obs. } (N) & \begin{bmatrix} \quad & \quad \\ \quad & \quad \end{bmatrix} \end{array}$$

$w_1 \quad \dots = \begin{bmatrix} w_{11} \\ w_{12} \\ w_{12} \\ \vdots \\ w_{1d} \end{bmatrix} \qquad w_0 = \begin{bmatrix} w_{01} \\ w_{02} \\ w_{03} \end{bmatrix}$

$\phi_1(x)$ "Linear regression"



$P(y=1 \mid x, w) = \sigma(\phi(x)^T w)$

Length d

$$X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \phi(x) = \underbrace{\begin{bmatrix} x^T w_1 \\ x^T w_2 \\ x^T w_3 \end{bmatrix}}_{n} = \begin{bmatrix} x_1 w_{11} + x_2 w_{12} \\ x_1 w_{21} + x_2 w_{22} \\ x_1 w_{31} + x_2 w_{32} \end{bmatrix} \cdot \begin{bmatrix} w_{01} \\ w_{02} \\ w_{03} \end{bmatrix}$$

a embedding

$$y = \phi(x)^T w_0$$

$$= w_{01}(x_1 w_{11} + x_2 w_{12}) + w_{02}(x_1 w_{21} + x_2 w_{22}) + w_{03}(x_1 w_{31} + x_2 w_{32})$$

$$= \underbrace{x_1 w_{11} w_{01} + x_1 w_{21} w_{02} + x_1 w_{31} w_{03}}_{x_1 \underbrace{(w_{11} w_{01} + w_{21} w_{02} + w_{31} w_{03})}_{w_1^*}} + \ldots$$

→ Linear in our Input d

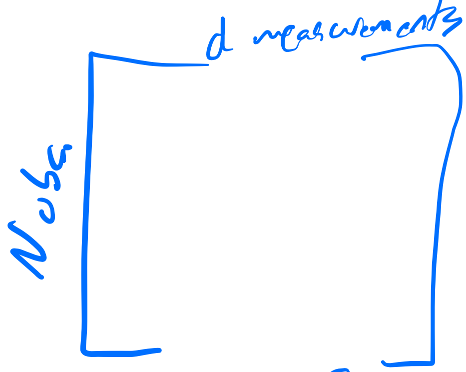
$$\underset{\sim}{W} = \begin{bmatrix} w_1^T \\ w_2^T \\ w_3^T \end{bmatrix} \xrightarrow{d} \begin{bmatrix} w_{11} & w_{12} & w_{13} & \ldots \\ w_{21} & w_{22} & w_{23} & \ldots \\ w_{31} & w_{32} & w_{33} & \ldots \end{bmatrix}$$

$$\phi(x) = \begin{bmatrix} \sigma(x^T w_1) \\ \sigma(x^T w_2) \\ \sigma(x^T w_3) \end{bmatrix} = \sigma(\underbrace{x^T W^T}_{(1 \times d)(d \times q) = (1 \times q)}) = \sigma(\underset{q \times d}{W} \underset{d}{x})$$

$$x = \begin{bmatrix} \; \end{bmatrix}$$

For a dataset $X = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix}$

X = $N \times d$ matrix

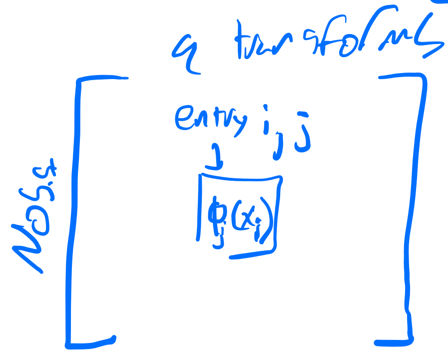


W = $q \times d$ matrix

$\phi(X) = \sigma(\underbrace{X W^T}_{\text{matrix}})$ ⟶

$(N \times d)(d \times q) \to N \times q$

np.dot



Regression

vector

$y = \phi(x) w_0 = \sigma(x^T W_1^T) w_0$

Logistic Reg.

$P(y=1 \mid x, W_1, w_0) = \sigma(\sigma(x^T W_1^T) w_0)$

Bias / offset

$y = x^T w + \boxed{b}$

$X \to \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \to \begin{bmatrix} x_1 \\ x_2 \\ 1 \end{bmatrix}$

$w \div \begin{bmatrix} w_1 \\ w_2 \end{bmatrix} \to \begin{bmatrix} w_1 \\ w_2 \\ b \end{bmatrix}$

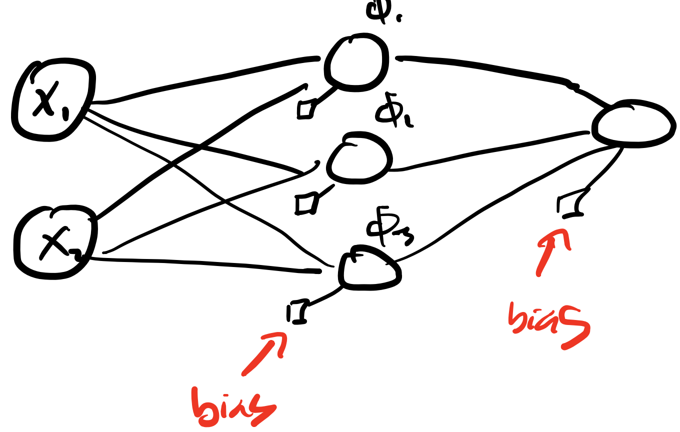
explicit offset

$\sigma(x^T W_1^T + b_1) w_0 + \boxed{b_0}$

Vector $(q \times 1)$

Scalar

$\phi(x) = \begin{bmatrix} \sigma(x^T v_1) + b_{11} \\ \sigma(x^T v_2) + b_{12} \\ \vdots \end{bmatrix}$



Learn parameters $\underbrace{W_1}$, w_0 [b_1 and b_0]

$\longrightarrow$ Gradient descent

$Loss(\bar{X}, y, W_1, w_0)$

Update $W_1^{(i+1)} \longleftarrow W_1^{(i)} - \alpha \nabla_{W_1} Loss(\bar{X}, y, W_1^{(i)}, w_0^{(i)})$

$w_0^{(i+1)} \longleftarrow w_0^{(i)} - \alpha \nabla_{w_0} Loss(\bar{X}, y, W_1^{(i)}, w_0)$

Shape of $\nabla_{W_1} Loss(\quad)$? $\longrightarrow$ $q \times d$

W_1 : $q \times d$ matrix

$$\nabla_{W_1} Loss = \begin{bmatrix} \frac{\partial L}{\partial W_{11}} & \frac{\partial L}{\partial W_{12}} \\ \frac{\partial L}{\partial W_{21}} & \frac{\partial L}{\partial W_{22}} \\ & \end{bmatrix}$$

Update (W_1, w_0)

$(W_1, w_0)^{(i+1)} \longleftarrow (W_1, w_0)^{(i)} - \alpha \nabla \quad Loss$