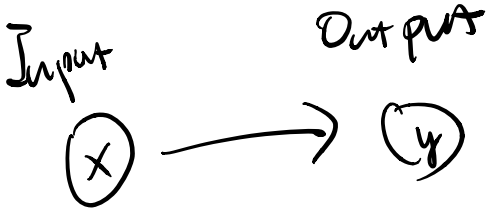
One-layer NN

activation function

$f(xw + b) = \hat{y}$



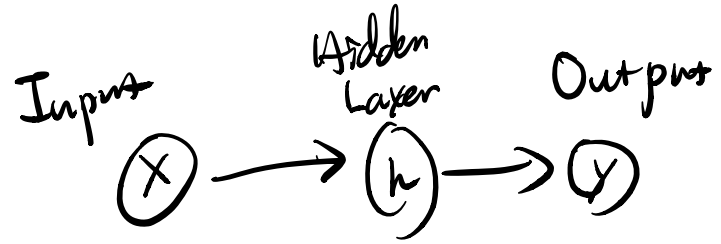
If $f(x) = x \Rightarrow \hat{y} = xw + b$ ← linear model

$\min_{b,w} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$ ← RSS → O.L.S.

Sigmoid: If $f(x) = \frac{1}{1+e^{-x}} \Rightarrow \hat{y} = \frac{e^{xw+b}}{1+e^{xw+b}}$ ← linear logistic regression

$\min_{b,w} -\sum_{i=1}^{n} \{ y_i \log \hat{y}_i + (1-y_i) \log (1-\hat{y}_i) \}$ ← Cross Entropy

Two-layer NN



$h = f_1(xw^{(1)} + b^{(1)})$

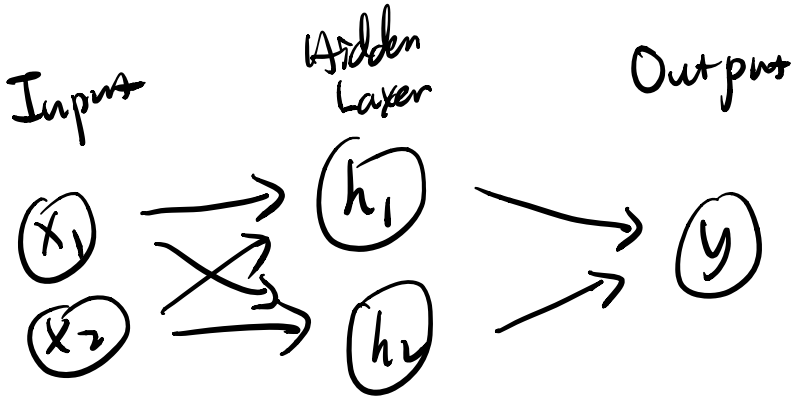
$y = f_2(hw^{(2)} + b^{(2)})$

plug h into y,

$y = f_2(f_1(xw^{(1)} + b^{(1)}) \cdot w^{(2)} + b^{(2)})$

Usually $f_1 = f_2$. Now, what if $f_1(x) = f_2(x) = x$?

More two-layer NNs



$$h_1 = f(x_1 w_{11}^{(1)} + x_2 w_{21}^{(1)} + b_1^{(1)})$$

$$h_2 = f(x_1 w_{12}^{(1)} + x_2 w_{22}^{(1)} + b_2^{(1)})$$

$$y = f(h_1 w_1^{(2)} + h_2 w_2^{(2)} + b^{(2)})$$

plug h_1, h_2 into y

$$y = f\Big(f(x_1 w_{11}^{(1)} + x_2 w_{21}^{(1)} + b_1^{(1)}) w_1^{(2)} + f(x_1 w_{12}^{(1)} + x_2 w_{22}^{(1)} + b_2^{(1)}) w_2^{(2)} + b^{(2)} \Big)$$