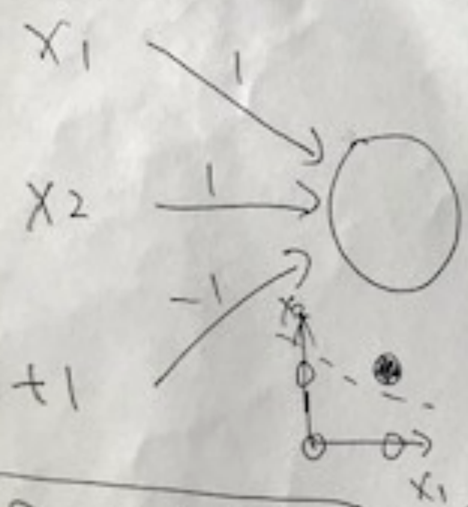
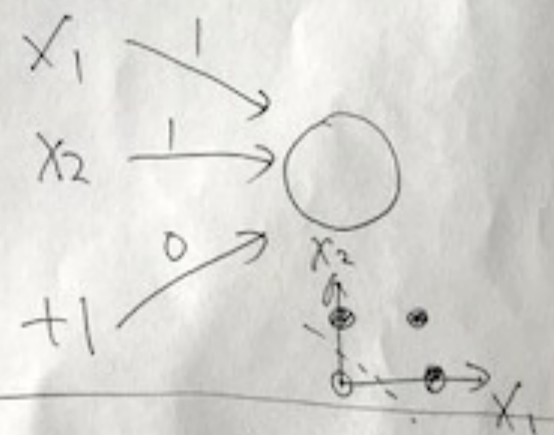
logical AND

x_1	x_2	y
0	0	0
0	1	0
1	0	0
1	1	1

OR

x_1	x_2	y
0	0	0
0	1	1
1	0	1
1	1	1





Perceptron : $y = \begin{cases} 0, & \text{if } wx + b \le 0 \\ 1, & \text{if } wx + b > 0 \end{cases}$

decision boundry -

$w_1x_1 + w_2x_2 + b = 0$

$x_2 = -(w_1/w_2)x_1 - b$

$\log \frac{1}{1+e^{-wx}}$

$= \log(1) - \log(1+e^{-wx})$

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

$\tanh(0) = 0$

$\tanh(1) \approx 0.76$ $\quad\tanh(-1) \approx -0.76$

$\tanh(2) \approx 0.96$

x= too many drug trials, too few patients

$y^* =$ health

$i^* = 0$

$L(x,y) = \log(0.21)$

$$\text{softmax}\left(\underbrace{[\overset{\text{health}^*}{6.05}, \overset{\text{sports}}{22.2}, \overset{\text{science}}{0.55}]}_{Wz}\right) = [\overset{*}{0.21}, 0.77, 0.02]$$

Logistic Regression

~~$L(x,y) = \log P(y^* \mid x)$~~

$$L(x,y) = \log P(y = y^* \mid x) = wf(x, y^*) - \log \sum_y e^{wf(x,y)}$$

logistic $(w^T x)$

gradient :

$$\frac{\partial L(x,y)}{\partial w_i} = x_i \left(y - \underline{P(y=+ \mid x)} \right)$$

$$= x_i (y - \underline{\hat{y}})$$

$$L(x, y=+) = \log P(y=+ \mid x) = \cancel{\log(\hat{y}) - \log(1+}$$

multi-class

$wx - \log(\sum e^{wx})$

$$= \boxed{\log \frac{e^{w^T x}}{1+e^{w^T x}}} = \log \frac{1}{1+e^{-wx}} = \log(1) - [\log(1+e^{-wx}$$

$= \cancel{\log(1) - [\log(1+e^{-\hat{y}}}$

logistic loss

0/1 Loss (actual classification error)

$y = \{-1, +1\}$

$$loss_{0/1} = \begin{cases} 0 & , y(wx+b) > 0 \\ 1 & , \text{otherwise} \end{cases}$$

Hinge Loss (Standard SVM)

$$loss_{Hinge} = \max(0, \underbrace{1 - y(wx+b)}_{1 - y\hat{y}})$$

$\hat{y} = wx + b$

ideas of surrogate loss (e.g. SVM)

- approximate 0/1 loss with a convex function
- surrogate losses are always upper bounds on the true loss function. This guarantees that minimizing surrogate loss → minimize true loss

Log-Loss (Logistic Regression)

$$Loss_{log} = \frac{1}{\ln 2} \boxed{\log(1 + e^{-y\hat{y}})}$$

$$loss_{log} = \frac{1}{\ln 2} \log(1 + e^{-y\hat{y}})$$

$\times(1 - \hat{y})$