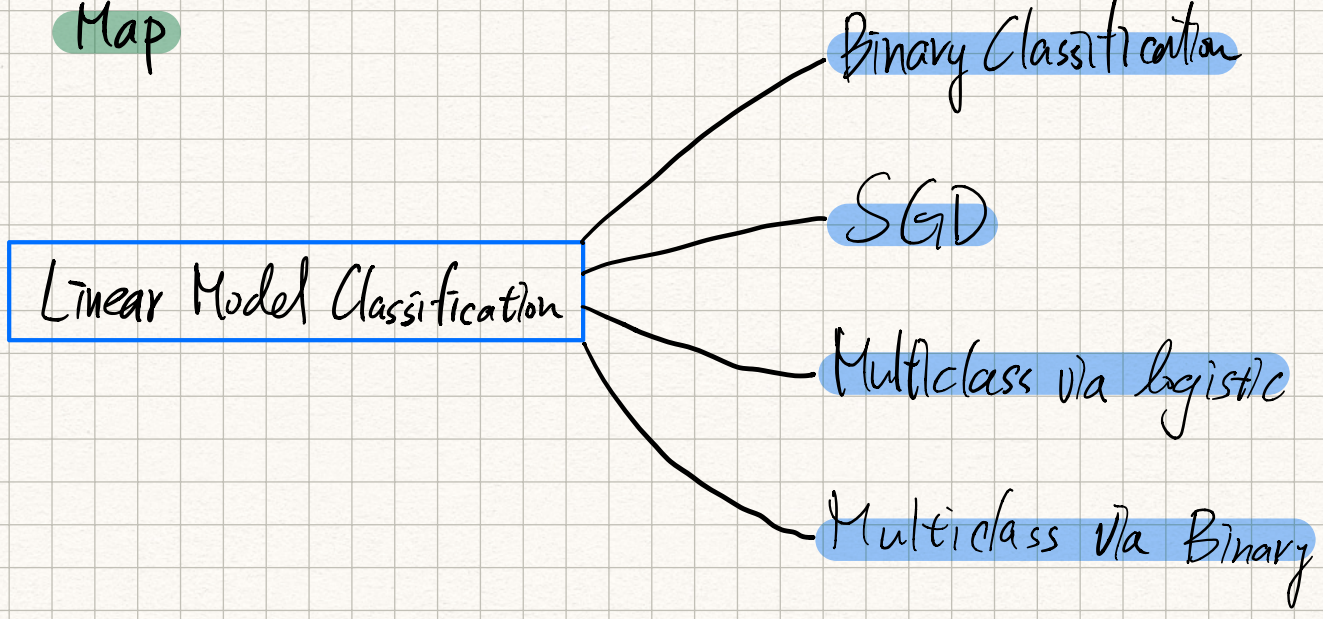
Map



Binary Classification

S: score，對所有特徵做線性組合所得的值

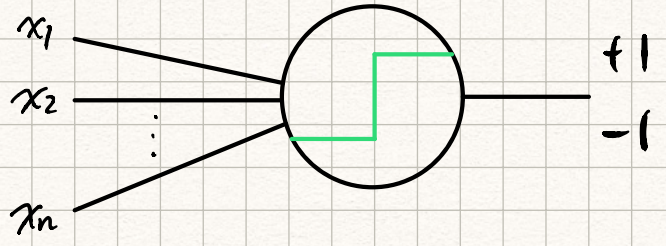
$$S = \sum_{i=1}^{n} w_i x_i = w^T x$$

h: hypothesis，在二維平面上一個 hypothesis 可視為一條線

而一個 h 可對應到一組 w

對 Binary classification 來說，h(x) 是對 S 取符號出來，

$$\Rightarrow h(x) = sign(s) = sign(w^T x)$$



$$err_{0/1}(w_i) = [\![h(x) \neq y]\!] = [\![sign(w^T x) \neq y]\!] = [\![sign(s) \neq y]\!]$$

$\min_w err_{0/1}$ → 最佳化求解不好解，故採 Pocket 的方式

一直做到夠久 ok 可接受為止

linear regression

直接求值

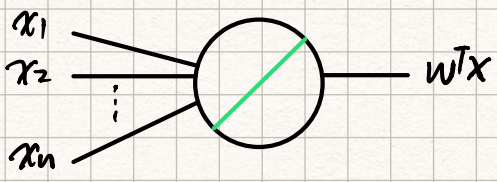
$$s = w^T x$$

$$h(x) = w^T x$$

$$err(w) = (h(x) - y)^2$$

$$= (s - y)^2$$

$$\min_w err \rightarrow w = X^+ y$$



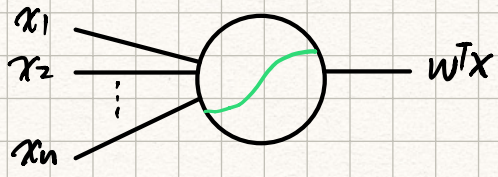
logistic regression

求机率

$$s = w^T x$$

$$\theta(s) = \frac{1}{1+e^{-s}}$$

$$h(x) = \theta(s) = \frac{1}{1+e^{-w^T x}}$$



$h(yx)$ $\ln(1+e^{-ys})$

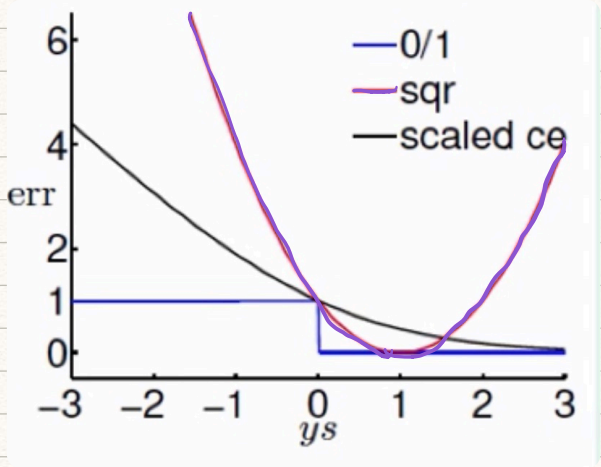
$$err(w) = \text{cross entropy} = -\ln h(x) = \ln(1+e^{-s})$$

$$\min_w err \rightarrow w_{t+1} \leftarrow w_t + y_n x_n$$

- 因为 binary classification 的 err 最佳化不好解，故想透过 linear regression to logistic regression 来协助求解。
- 要透过 linear regression to logistic regression 来协助求解，就需把 err_{01} 跟 err_{sqr} or err_{log} 连结起来

1. $err_{01} = [\![sign(s) \neq y]\!] = [\![sign(ys) \neq 1]\!]$
2. $err_{sqr} = (s-y)^2 = (ys-1)^2$
3. $err_{sce} = -y \ln(h(x)) = -\ln(h(ys))$

特上述3條式子圖解：



$err_{sce} = \frac{1}{\ln 2} err_{ce}$

- 從圖中可以得知 $err_{0/1} \le err_{sqr}$, $err_{0/1} \le err_{sce}$，這就表示：只要把 err_{sqr} 或 err_{sce} 做小，$err_{0/1}$ 也就會隨之下降

SGD

什麼情況下需要 SGD？想加快更新速度

在做 GD 時總是要把所有的資料跑完才可以得到更新值

$$w_{t+1} \leftarrow w_t + \frac{1}{N}\sum_{n=1}^{N} \theta(s) y_n x_n$$

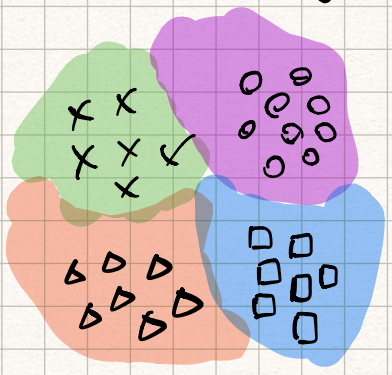
$\theta(-y_n w^T x_n)$

$-\nabla E_{in}(w)$

當資料量大時更新速度會變慢，故以PLA為參考修改為一次更新一個點

Multiclass via logistic

OVA ≡ One Verse All



$x=[x_1, x_2]$

$s_\square = w_\square^T x$ → $P(y=\square|x)$

$s_\times = w_\times^T x$ → $P(y=\times|x)$

$s_\circ = w_\circ^T x$ → $P(y=\circ|x)$

$s_\triangle = w_\triangle^T x$ → $P(y=\triangle|x)$

θ → max P

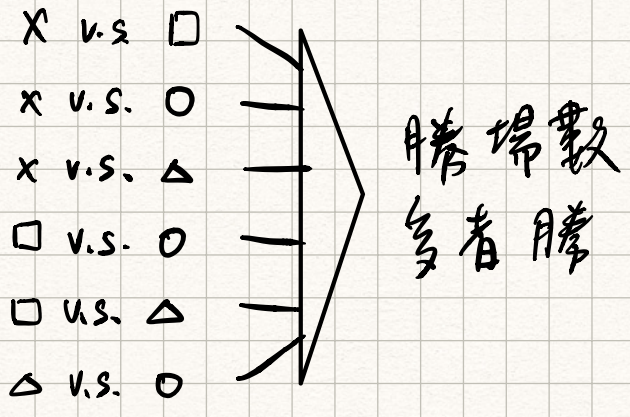
可能會遇到資料不平衡的問題

正的資料遠少於其他

Multiclass Via Binary

OVO ≡ One Verse One

概念有點像打比賽：C_2^4



Q：若有2隊有相同勝場數怎麼辦？