1.

$$\max \quad 1 - \sum^{K} \mu_K^2$$

$$\text{s.t} \quad \sum^{K} \mu_K = 1$$

$$L = 1 - \sum^{K} \mu_K^2 + \lambda \left(\sum^{K} \mu_K - 1 \right)$$

$$\forall k \quad \frac{\partial L}{\partial \mu_k} = -2\mu_k + \lambda \mu_k \Rightarrow \mu_k = \frac{\lambda}{2}$$

$$\sum^{K} \mu_k = 1 \Rightarrow K \cdot \frac{\lambda}{2} = 1 \Rightarrow \lambda = \frac{2}{K}.$$

$$\mu_K^* = \frac{1}{K}$$

$$\text{max impurity} = 1 - K \cdot \frac{1}{K^2} = 1 - \frac{1}{K}$$

2.

$$\mu_+ \left(1 - 2(\mu_+ - \mu_-) + (\mu_+ - \mu_-)^2 \right)$$

$$+ \mu_- \left(1 - 2(\mu_+ - \mu_-) + (\mu_+ - \mu_-)^2 \right)$$

$$= 1 + (\mu_+ - \mu_-)^2 - 2\mu_+^2 - 2\mu_-^2 + 4\mu_+\mu_-$$

$$= 1 - \mu_+^2 - \mu_-^2 \ 2\mu_+\mu_- \quad \neq r(1 - \mu_+^2 - \mu_-^2)$$

$$\downarrow r \in \mathbb{R}.$$

$\Rightarrow$ not scaled Gini impurity.

3.

$$\lim_{N\to\infty}\left(1-\frac{1}{N}\right)^{pN} = \left[\left(1-\frac{1}{N}\right)^{N}\right]^{p} \approx e^{-p}$$

↳ 抽出 N' 個後，未被抽到的比例

$e^{-p}\cdot N$ is the number of unsample

4.

K個 classification.

至少要 $\frac{K+1}{2}$ 個都犯錯 G 才會出錯

在極端的情況下，將所有的錯誤 $\sum^{K} e_k$

集中在 $\frac{K+1}{2}$ 個分類器上，會使 $E_{out}(G)$ 放大.

此時 $E_{out}(G) \le \sum^{K} e_k \Big/ \frac{K+1}{2} = \frac{2}{K+1}\sum^{K} e_k$

5. By Lecture 11 P.17.

α_1 is optimal η by g_1-transformed LinearReg

which $\min_{\eta} \frac{1}{N}\sum\left((y_n-\underbrace{s_n}_{0})-\eta\underbrace{g_1(x_n)}_{11.26}\right)^2$

By Formula of OLS.

$$\alpha_1=\eta=\frac{\sum^{N} g_1(x_n)(y_n-\overset{0}{s_n})}{\sum^{N}\left[g_1(x_n)\right]^2} = \frac{1}{11.26}\;\frac{\sum^{N} y_n}{N}.$$

6.

$$\alpha_t = \overset{\eta \text{ in } t \text{ iteration}}{\eta} = \frac{\sum g_t(x_n)(y_n - s_n^{t-1})}{\sum g_t^2(x_n)}$$

$$\Rightarrow \alpha_t \sum g_t^2(x_n) = \sum g_t(x_n) \cdot y_n - \sum g_t(x_n) s_n^{t-1}$$

$$\Rightarrow \sum g_t \underbrace{(\alpha_t \sum g_t + s_n^{n-1})}_{s_n^t} = \sum g_t(x_n) \cdot y_n$$

$$\Rightarrow \sum^N s_n^t g_t(x_n) = \sum^N g_t(x_n) \cdot y_n$$

7. ~~By~~ Lecture 11. P19.

initial $= 0$.

$$\cancel{\alpha_1 = \eta^* = \frac{\sum^N g_t(x_n)(y_n - s_n^0)}{\sum^N g_t^2(x_n)}}$$

$$\cancel{= \frac{\sum^N g_t(x_n) \cdot y_n}{\sum^N g_t^2(x_n)}}$$

By Problem 6 $= \cancel{\dfrac{\sum s_n^1 \cdot g_t(x_n)}{\sum^N g_t^2(x_n)} = \dfrac{\sum g_t^2(x_n)}{\sum g_t^2(x_n)}} = 1$

$(s_n^1 = g_1(x_n))$

7.

Initially $s_1 = \cdots = s_n = 0$

In the first iteration:

we find gt by running squared error regression on $\{(x_n, y_n)\}$

after find gt

when we find α_1

since we have to minimize $\min_{\eta} \frac{1}{N} \sum (y_n - \eta g_t(x_n))^2$

$\alpha_1 = \eta$ must be 1, since its the same objection fuction of regression in the first ~~iterat~~ regression.

8. OR. $x_1 \dots x_d$ 只有為 -1 才輸出 -1
otherwise +1

$(w_0, w_1, \dots w_d)$

$= (d-1, 1, \dots, 1)$

when $x_i \ \forall i = -1$.

$\sum_i^d w_i x_i = -1 \Rightarrow g_A(x) = -1$

9. For output layer.

$$\frac{\partial e_n}{\partial w_{i1}^{(L)}} = -2(y_n - s_1^{(L)}) \cdot (x_i^{(L-1)})$$

For other layer

$$\frac{\partial e_n}{w s_j^{(\ell)}} = \delta^{(\ell)} \cdot (x_1^{\ell-1})$$

By back prop : $\delta_j^{(\ell)} = \sum_k \delta_k^{(\ell+1)} (w_{jk}^{\ell+1}) \tanh'(s_j^{(\ell)})$

Since $w_{ij}^{(\ell)} = 0 \Rightarrow \delta_j^{(\ell)} = 0 \quad \forall \ \ell < L, j$

only $x_0^{L-1} = 1$, $w_i^{L-1} = 0$

$\Rightarrow$ only $\frac{\partial e_n}{\partial w_{01}^{(L)}}$ may not be zero.

10.

$$\frac{\partial e}{\partial s_k^{(L)}} = \frac{\partial -\sum v_k \ln q_k}{\partial s_k}$$

$$= \frac{\partial -(v_1 \ln q_1 + v_2 \ln q_2 + \cdots + v_k \ln q_k + \cdots)}{\partial s_k}$$

$$= -\left(v_1 \frac{1}{q_1}\frac{\partial q_1}{\partial s_k} + \cdots + v_k \frac{1}{q_k}\frac{\partial q_k}{\partial s_k} \quad + v_K \frac{1}{q_K}\frac{\partial q_K}{\partial s_K}\right)$$

$$\begin{cases} \dfrac{\partial q_i}{\partial s_j} = -q_i q_j \quad \text{for } i \neq j \\ \dfrac{\partial q_i}{\partial q_i} = q_i(1-q_i) \end{cases}$$

$$= -\left(v_1 \frac{1}{\cancel{q_1}} \cancel{q_1} q_k + \cdots + v_k \frac{q_k}{q_k}(1-q_k) + \cdots\right)$$

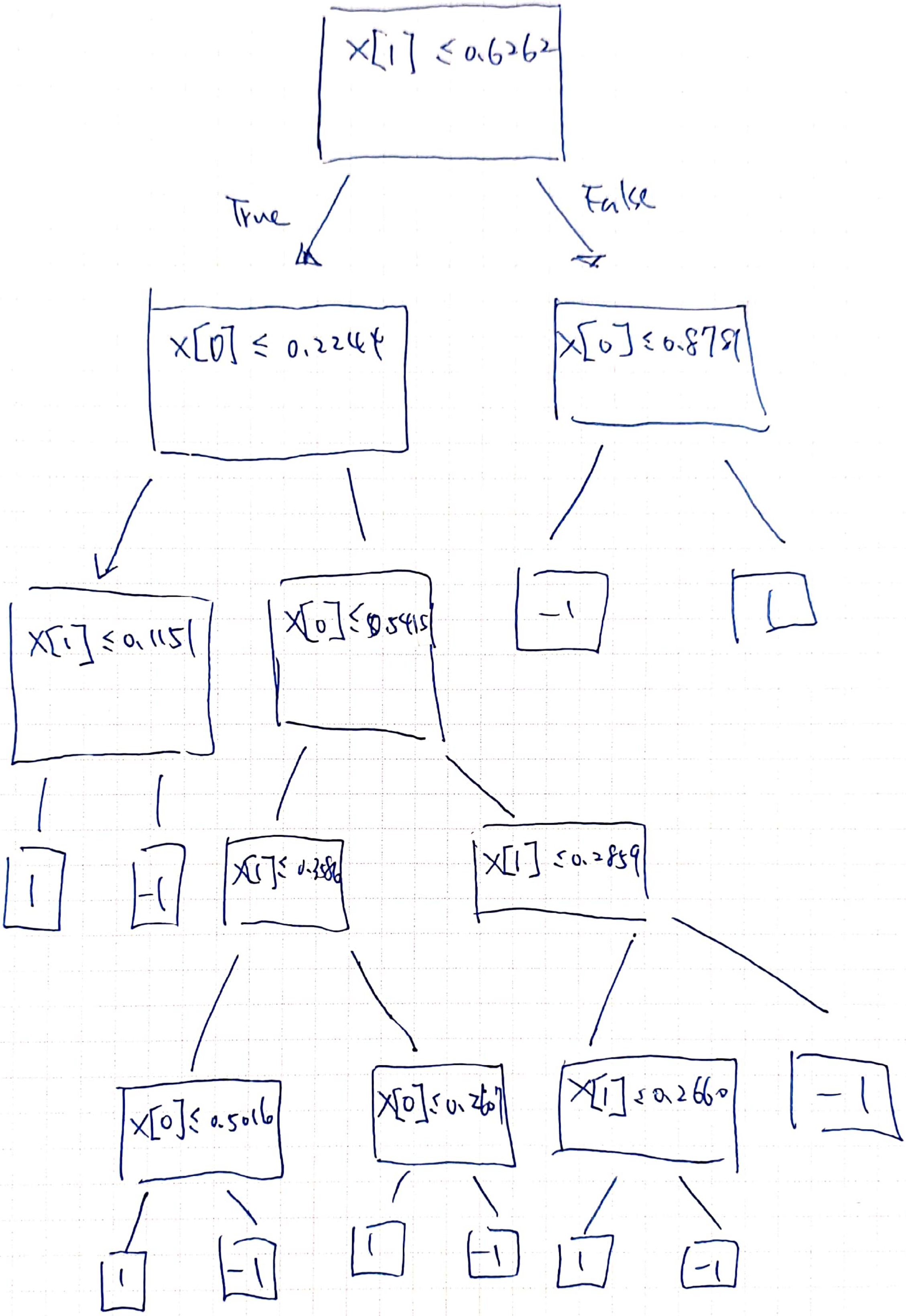
$$= -(-v_1 q_k - \cdots \quad - v_k q_k + v_k + \cdots)$$

since for $v_1 \ldots v_k \ldots v_K$, only one will be 1

others are 0

$$= -(-q_k + v_k) = v_k - q_k. \#$$

11.



X[1] ≤ 0.6262
True
False
X[0] ≤ 0.2244
X[0] ≤ 0.8789
X[1] ≤ 0.1151
X[0] ≤ 0.5415
-1
1
1
-1
X[1] ≤ 0.3886
X[1] ≤ 0.2859
X[0] ≤ 0.5016
X[0] ≤ 0.2607
X[1] ≤ 0.2660
-1
1
-1
1
-1
1
-1