FAI HW3

1 Find gradeint direction for E_{in}

$E_{in}(w) = \frac{1}{N}\sum_{n=1}^{N}(max(1-y_n w^T x_n, 0))^2$
$\nabla E_{in}(w) = \frac{\partial E_{in}(w)}{\partial w}$
$= \frac{\partial max(1-y_n w^T x_n, 0)^2}{\partial f(x^n)} \frac{\partial f(x^n)}{\partial w_i} x_i^n$
$\begin{cases} 0, \text{ if } 1-y_n w^T x_n < 0 \\ -2(1-y_n), \text{ otherwise} \end{cases}$
$\nabla E_{in}(w) = \frac{1}{N}\sum_{n=1}^{N} -2max((1-y_n)w^T, 0)x_n$

2 Prove $w_{lin} = V\Gamma U^T y$

The linear regression solution $w_{lin} = (X^T X)^{-1} X^T y$, and need to prove the general solution $w_{lin} = V\Gamma U^T y$ is the same as showing $V\Gamma U^T$ regardless of whether $X^T X$ is invertible.

$A = U \sum V^T$
$A^{\dagger} = V \sum^{\dagger} U^T$
proof by verifying the fout conditions
$AA^{\dagger}A = U \sum V^T * V \sum^{\dagger} U^T * U \sum V^T = A$
$A^{\dagger}AA^{\dagger} = V \sum^{\dagger} U^T * U \sum V^T * V \sum^{\dagger} U^T = A^{\dagger}$
$AA^{\dagger} = U \sum V^T * V \sum^{\dagger} U^T = U \sum \sum^{\dagger} U^T$ symmetric
$A^{\dagger}A = V \sum^{\dagger} U^T * U \sum V^T = V \sum^{\dagger} \sum V^T$ symmetric
another approach
$w_{lin} = (X^TX)^{-1}X^Ty$
$= [(V \sum^T U^T)(U \sum V^T)]^{-1}(V \sum^T U^T)y$
$= V \sum^{-2} V^TV \sum^T U^Ty$
$= V \sum^{-1} U^Ty$
$= V\Gamma U^Ty$
since $\Gamma[i, j] = \frac{1}{\sum[j,i]} if \sum[j, i] > 0 and \Gamma[i, j] = 0$ otherwise
so the general solution is $w_{lin} = V\Gamma U^Ty$

3

Reference:

[https://people.missouristate.edu/songfengzheng/Teaching/MTH541/Lecture notes/MLE.pdf](https://people.missouristate.edu/songfengzheng/Teaching/MTH541/Lecture%20notes/MLE.pdf)

(https://people.missouristate.edu/songfengzheng/Teaching/MTH541/Lecture%20notes/MLE.pdf),

B09902029

let

$$L(u,\sum) = ln\Pi_{n=1}p_u(x_n)$$

$$= ln\Pi_{n=1}^{N} - \frac{1}{2}ln(2\pi) - \frac{1}{2}ln(|I|) - \frac{1}{2}(X_n-u)^T I^{-1}(X_n-u)$$

$$\frac{\partial L(u,\sum)}{\partial \mu} = 0 \rightarrow \sum_{n=1}^{N}(x_n-u) = 0 \rightarrow u* = argmax\pi_{n=1}^{N}L(u,\sum) = \frac{1}{N}\sum_{n=1}^{N}x_n$$

4

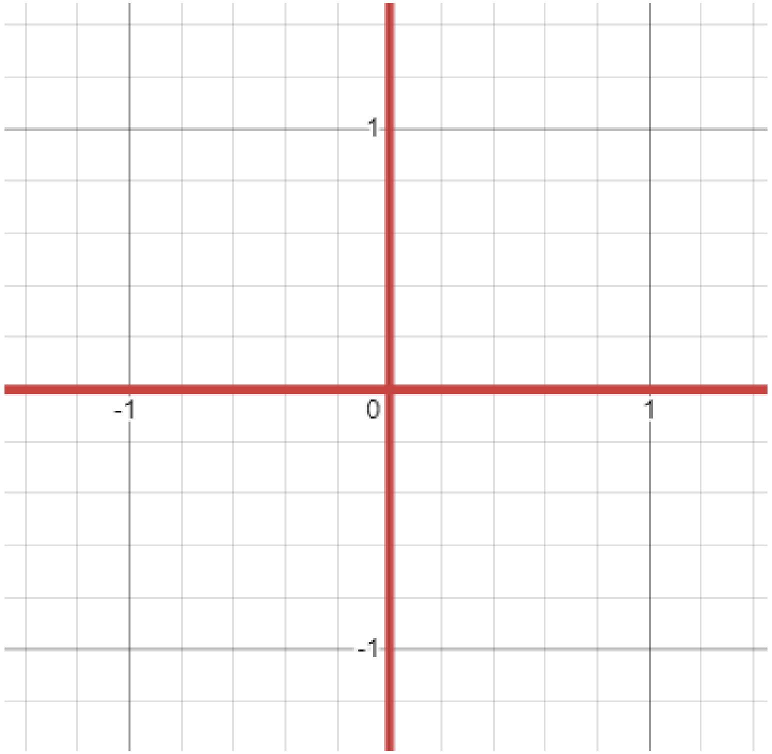
$\mathbf{z} = \Phi_2(\mathbf{x})$	y
$\mathbf{z}_1 = \Phi_2(\mathbf{x}_1)$	$y_1 = -1$
$\mathbf{z}_2 = \Phi_2(\mathbf{x}_2)$	$y_2 = +1$
$\mathbf{z}_3 = \Phi_2(\mathbf{x}_3)$	$y_3 = -1$
$\mathbf{z}_4 = \Phi_2(\mathbf{x}_4)$	$y_4 = +1$

$z = \Phi_2(x)$	**y**
$z_1 = \Phi_2([+1, +1])$	$y_1 = -1$
$z_2 = \Phi_2([-1, +1])$	$y_2 = +1$
$z_3 = \Phi_2([-1, -1])$	$y_3 = -1$
$z_4 = \Phi_2([+1, -1])$	$y_4 = +1$

$y_n = sign(\hat{w}^T x_n) for \mathrm{n} = 1, 2, 3, 4$

say $w = x_1 x_2$

plot $\hat{w}^T \Phi_2(x) = 0$



Programming report

用f1 score 作為好壞的原因是因為我們想要找出 spam 的 email,所以如果spam被找出來(tp)才有意義

8

acc, f1

	Logistic Regression	Decision Tree	Random Forest
1	0.78125, 0.46153846	0.78125, 0.36363636	0.78125, 0.22222222
2	0.875, 0.75	0.71875, 0.30769230	0.71875, 0.181818182
3	0.84375, 0.73684210	0.6875, 0.5	0.65625, 0.0
4	0.9375, 0.88888888	0.78125, 0.58823529	0.6875, 0.28571428
5	0.90625, 0.84210526	0.5, 0.27272727	0.65625, 0.266666666
best	0.9375, 0.88888888	0.78125, 0.58823529	0.6875, 0.28571428

9

5	11	17
0.22222222	0.166666666	0.2
0.181818181	0.2	0.2
0.0	0.166666666	0.166666666
0.28571428	0.461538461	0.461538461
0.266666666	0.25	0.235294117

11 棵樹的效果最好