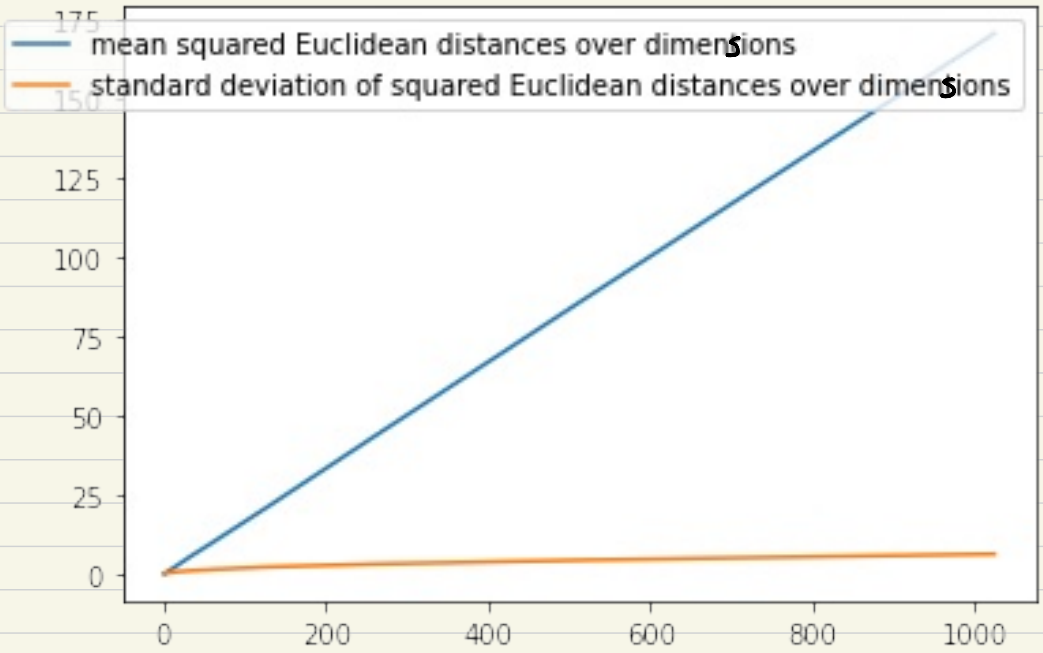
HW1

1) a)



b)

$R = Z_1 + \dots + Z_d \quad , \quad Z_i = (X_i - Y_i)^2$

$X = \begin{bmatrix} X_1 \\ \vdots \\ X_d \end{bmatrix}$ $Y = \begin{bmatrix} Y_1 \\ \vdots \\ Y_d \end{bmatrix}$

$E[R] = E[Z_1 + \dots + Z_d]$

$= E[Z_1] + \dots + E[Z_d]$

$= d E[Z_i] = \frac{d}{6}$

$Var[R] = Var[Z_1 + \dots + Z_d]$

$= Var[Z_1] + \dots + Var[Z_d]$ since Z_i's are independent

$= d \frac{7}{180}$

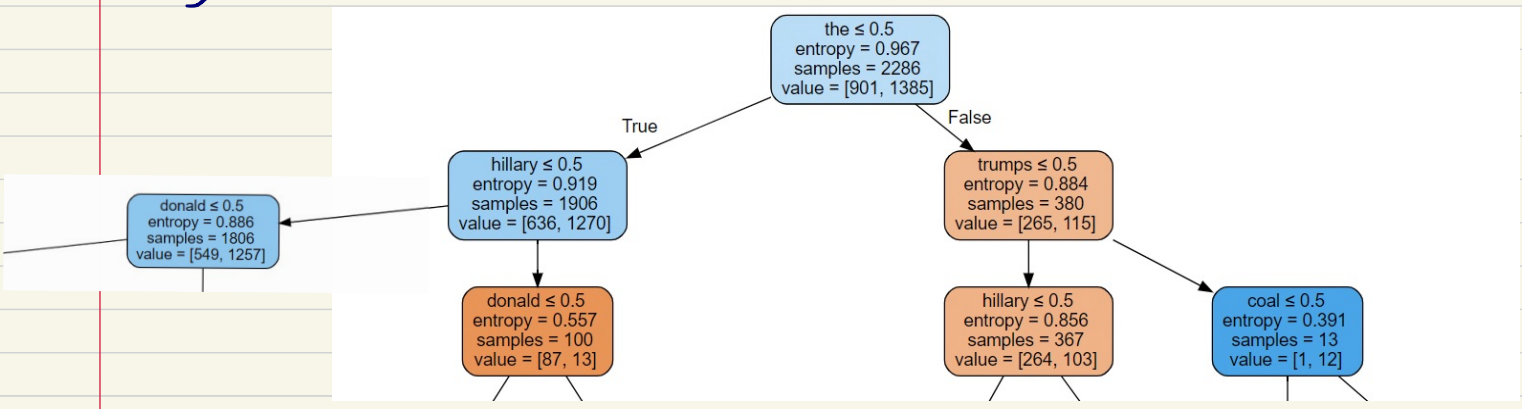
2) a) see submitted code.

b) performance of model with various max_depth and split_criteria.

	max_depth	split_criteria	score
0	31	entropy	0.773469
1	41	gini	0.771429
2	46	gini	0.769388
3	16	gini	0.765306
4	36	gini	0.765306
5	46	entropy	0.763265
6	36	entropy	0.761224
7	41	entropy	0.759184
8	26	entropy	0.757143
9	31	gini	0.757143
10	26	gini	0.755102
11	21	entropy	0.753061
12	21	gini	0.742857
13	11	gini	0.742857
14	16	entropy	0.738776
15	6	gini	0.716327
16	11	entropy	0.714286
17	6	entropy	0.700000
18	1	entropy	0.661224
19	1	gini	0.661224

The best model with 31 max depth and split-criteria 'entropy' had a 77.3% accuracy tested on validation set.

c)



d)

The top-most split is 'the'.

Information Gain of 'the' : 0.05445

Information Gain of 'hillary' : 0.043199...

Information Gain of 'trumps' : 0.04057...

Information Gain of 'donald' : 0.04694...

3)

a) $L = \frac{1}{2}(y-t)^2$

$$J(w,b) = \frac{1}{2N}\sum_{i=1}^{N}(y^{(i)}-t^{(i)})^2 + \frac{1}{2}\sum_{j=1}^{D}B_j - w_j^2$$

$$= \frac{1}{2N}\sum_{j=1}^{D}\left(\left[(w_j x_j^{(i)}) + b\right] - t^{(i)}\right)^2 + \frac{1}{2}\sum_{j=1}^{D}B_j - w_j^2$$

$$b \leftarrow b - \alpha\frac{\partial J}{\partial b} = b - \frac{\alpha}{N}\sum_{i=1}^{N}(y^{(i)}-t^{(i)})$$

so $\boxed{b \leftarrow b - \frac{\alpha}{N}\sum_{i=1}^{N}(y^{(i)}-t^{(i)})}$

$$w_j \leftarrow w_j - \alpha\frac{\partial J}{\partial w_j}$$

$$= w_j - \alpha\left[\frac{1}{2N}\frac{\partial}{\partial w_j}\left(\sum_{i=1}^{N}\left[\sum_{j=1}^{D}(w_j x_j^{(i)}) + b - t^{(i)}\right]^2 + \frac{\partial}{\partial w_j}\left(\frac{1}{2}\sum_{j=1}^{D}B_j w_j^2\right)\right)\right]$$

$$= w_j - \alpha\left[\left(\frac{1}{2N}\sum_{i=1}^{N}2\left(\sum_{j=1}^{D}(w_j x_j^{(i)}) + b - t^{(i)}\right)(x_j^{(i)})\right) + \left(\frac{1}{2}B_j 2w_j\right)\right]$$

$$= w_j - \alpha\left[\frac{1}{N}\sum_{i=1}^{N}(y^{(i)}-t^{(i)})x_j^{(i)} + B_j w_j\right]$$

$$= w_j - \frac{\alpha}{N}\sum_{i=1}^{N}(y^{(i)}-t^{(i)})x_j^{(i)} - \alpha B_j w_j$$

$$= (1-\alpha B_j)(w_j) - \frac{\alpha}{N}\sum_{i=1}^{N}(y^{(i)}-t^{(i)})$$

so $\boxed{w_j \leftarrow (1-\alpha B_j)(w_j) - \frac{\alpha}{N}\sum_{i=1}^{N}(y^{(i)}-t^{(i)})}$

so at each step, we update w_j based on the amount by which the estimates

$y^{(i)}$ differ from the target, however, we also try to decrease w_j by $\alpha B_j w_j$ each time towards 0, hence 'weight decay'.

b) $$\frac{\partial J_{reg}^{B}}{\partial w_j} = \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{j'=1}^{D}\left(w_{j'} x_{j'}^{(i)}\right) - t^{(i)}\right)\left(x_j^{(i)}\right) + B_j w_j$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left[\left(x_j^{(i)}\sum_{j'=1}^{D}\left(w_{j'} x_{j'}^{(i)}\right)\right) - x_j^{(i)} t^{(i)}\right] + \left(B_j w_j\right)$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left[\sum_{j'=1}^{D} w_{j'} x_{j'}^{(i)} x_j^{(i)} - x_j^{(i)} t^{(i)}\right] + \left(B_j w_j\right)$$

$$= \sum_{j'=1}^{D}\left[\underbrace{\frac{1}{N}\sum_{i=1}^{N}\left(x_j^{(i)} x_{j'}^{(i)}\right) + \mathbb{I}(j'=j) B_j}_{A_{jj'}}\right] w_{j'} \underbrace{- \frac{1}{N}\sum_{i=1}^{N} x_j^{(i)} t^{(i)}}_{\vec{c}_j}$$

so $$A_{jj'} = \frac{1}{N}\sum_{i=1}^{N}\left(x_j^{(i)} x_{j'}^{(i)}\right) + \mathbb{I}(j'=j) B_j$$

where $\mathbb{I}$ is the indicator function that returns 1 when $j'=j$ and 0 otherwise.

and $$\vec{c}_j = \frac{1}{N}\sum_{i=1}^{N} x_j^{(i)} t^{(i)}$$

c) $\underbrace{\left(\frac{1}{N}X^T X + \text{Diag}(B)\right)}_{A}\vec{w} - \underbrace{\frac{1}{N}X^T\vec{t}}_{\vec{c}}$

$\overset{\text{set}}{=} 0$

assuming that A is invertible,

$\Rightarrow A\vec{w} = \vec{c}$

$\Rightarrow \vec{w} = A^{-1}\vec{c}$

$\Rightarrow \boxed{\vec{w} = \left(\frac{1}{N}X^T X + \text{Diag}(B)\right)^{-1}\frac{1}{N}X^T\vec{t}}$