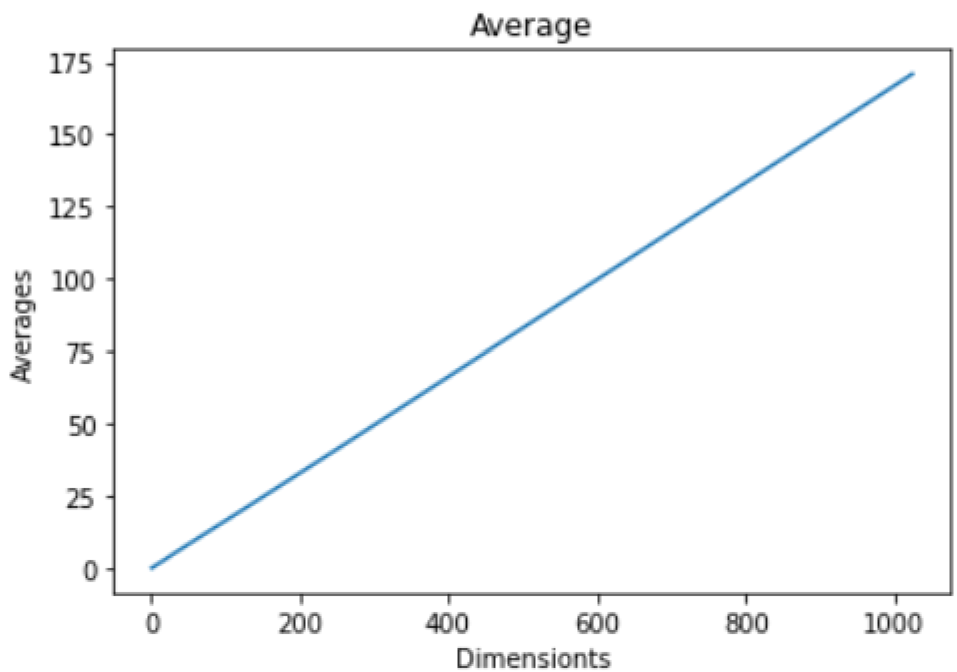
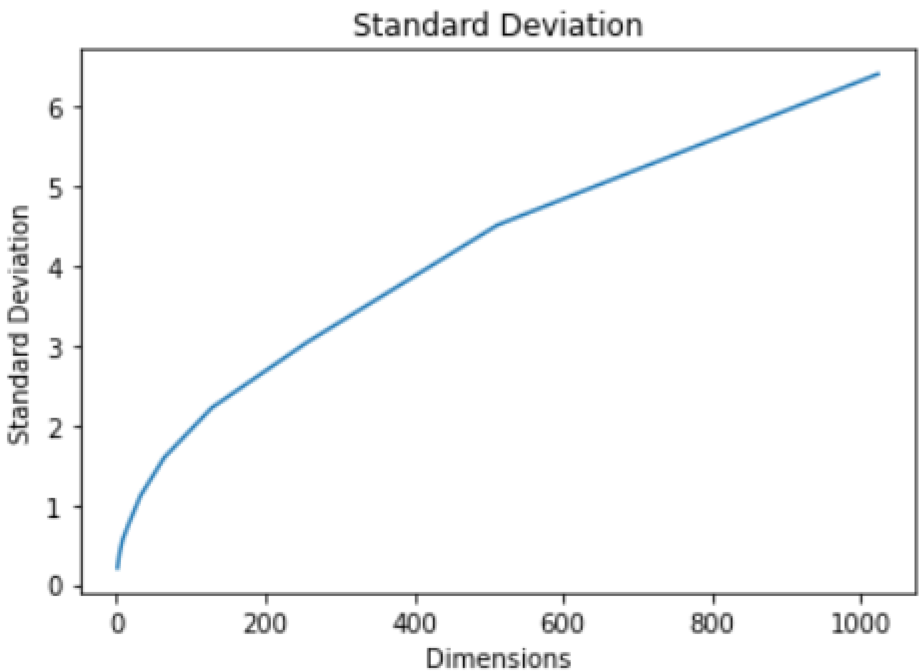
All the work that is specifically said to be included in hw1_writeup.pdf

Q1a: graphs of averages and standard deviations





Q1b: Calculate E[R] and var[R]

Q1b) ① $E[R] = E[z_1 + \cdots + z_d]$

$= E[z_1 + (z_2 + \cdots + z_d)]$

because $E[z_i + z_j] = E[z_i] + E[z_j]$

$= E[z_1] + E[z_2 + \cdots + z_d]$

$= E[z_1] + \cdots + E[z_d]$

$= d \cdot \frac{1}{6}$ # because $E[z_i] = \frac{1}{6}$

$= \frac{d}{6}$ #

② $V[R] = V[z_1 + \cdots + z_d]$

$= V[z_1 + (z_2 + \cdots + z_d)]$

$= V[z_1] + V[z_2 + \cdots + z_d]$

because if z_i and z_j are independent, then $Var[z_i + z_j] = Var[z_i] + Var[z_j]$

$= V[z_1] + \cdots + V[z_d]$

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

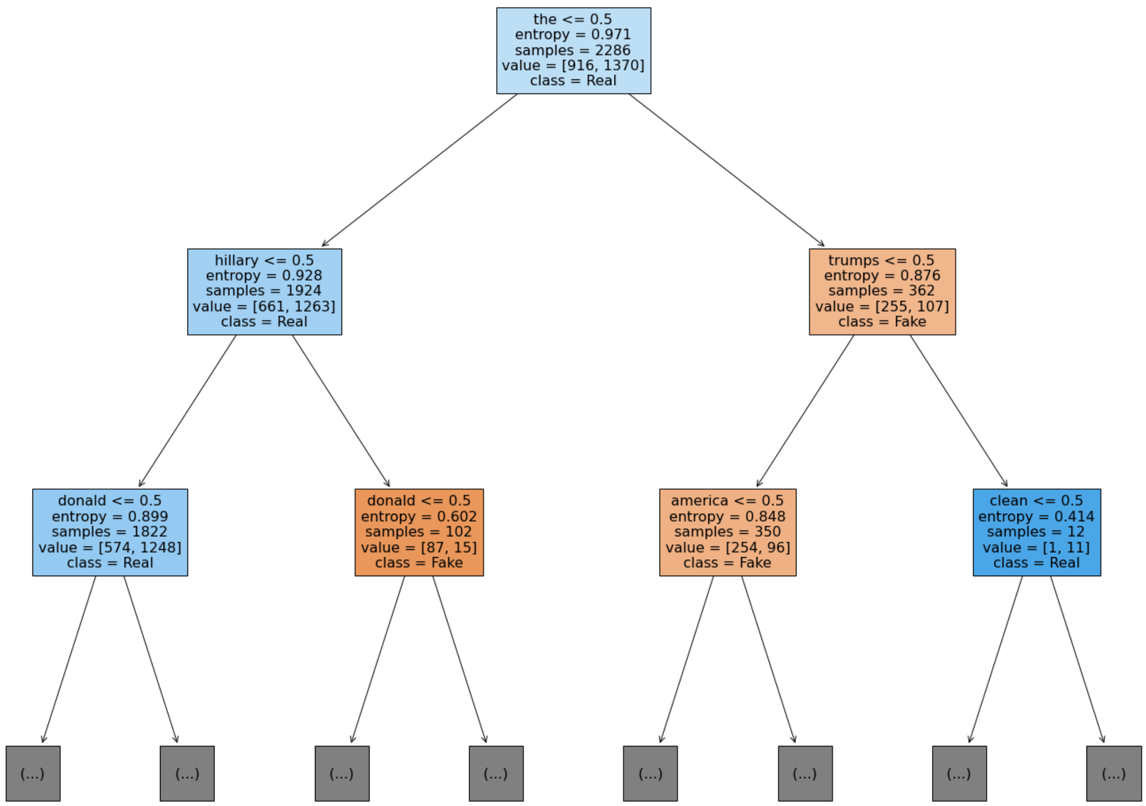
$= \frac{7d}{180}$

Hence $\begin{cases} E[R] = \frac{d}{6} \\ var[R] = \frac{7d}{180} \end{cases}$

Q2b: Prints the resulting accuracies of each model

```
[{'max_depth': 1, 'criterion': 'gini', 'accuracy': 0.6653061224489796},
 {'max_depth': 1, 'criterion': 'entropy', 'accuracy': 0.6653061224489796},
 {'max_depth': 2, 'criterion': 'gini', 'accuracy': 0.7183673469387755},
 {'max_depth': 2, 'criterion': 'entropy', 'accuracy': 0.7183673469387755},
 {'max_depth': 4, 'criterion': 'gini', 'accuracy': 0.7204081632653061},
 {'max_depth': 4, 'criterion': 'entropy', 'accuracy': 0.7204081632653061},
 {'max_depth': 8, 'criterion': 'gini', 'accuracy': 0.7204081632653061},
 {'max_depth': 8, 'criterion': 'entropy', 'accuracy': 0.7224489795918367},
 {'max_depth': 16, 'criterion': 'gini', 'accuracy': 0.7428571428571429},
 {'max_depth': 16, 'criterion': 'entropy', 'accuracy': 0.7591836734693878}]
```

Q2c: Tree of the model that archives the highest accuracy



Q2d: Information gain of different models

```
** Topmost **
Topmost(the) Information Gain: 0.0515
** Five more keywords **
['Split with trump Information Gain: 0.0325']
['Split with hillary Information Gain: 0.0373']
['Split with great Information Gain: 0.0001']
['Split with fake Information Gain: 0.0']
['Split with america Information Gain: 0.0096']
** Five more randomly picked keywords **
Split with chants Information Gain: 0.0017
Split with image Information Gain: 0.0
Split with international Information Gain: 0.0
Split with biggest Information Gain: 0.0003
Split with suddenly Information Gain: 0.0003
Split with independence Information Gain: 0.0006
Split with death Information Gain: 0.0012
Split with syrian Information Gain: 0.0
Split with disturbance Information Gain: 0.0
Split with semitism Information Gain: 0.0
```

Q3a

Q3a) First of all, let me calculate the following, which I'll reuse in part b.

$$\frac{\partial J_{reg}^{\beta}(W_j)}{\partial W_j} = \frac{\partial}{\partial W_j}\left[\frac{1}{2N}\sum_{i=1}^{N}(y^i - t^i)^2 + \frac{1}{2}\sum_{K=1}^{D}\beta_K W_K^2\right]$$

$$= \frac{1}{2N}\cdot\sum_{i=1}^{N} 2(y^i - t^i)\cdot(y^i - t^i)' + \frac{\partial}{\partial W_j}\frac{1}{2}\sum_{K=1}^{D}\beta_K W_K^2$$

$$= \frac{1}{2N}\cdot\sum_{i=1}^{N} 2(y^i - t^i)\cdot x_j^i + \frac{\partial}{\partial W_j}\frac{1}{2}\beta_1 W_1^2 + \cdots + \frac{\partial}{\partial W_j}\cdot\frac{1}{2}\beta_j W_j^2 + \cdots + \frac{\partial}{\partial W_j}\cdot\frac{1}{2}\beta_D W_D^2$$

because

$$\frac{\partial(y^i - t^i)}{\partial W_j} = \frac{\partial}{\partial W_j}\sum_{K=1}^{D} W_K x_K^i + b$$

$$= \frac{\partial}{\partial W_j}(\underbrace{W_1 x_1^i}_{0} + \cdots + W_j x_j^i + \cdots + \underbrace{W_D x_D^i}_{0} + b)$$

$$= x_j^i$$

$$= \frac{1}{N}\sum_{i=1}^{N}(y^i - t^i)\cdot x_j^i + \beta_j W_j \qquad ①$$

① And the update rule for w_j is:

$$w_j \leftarrow w_j - a \cdot \frac{\partial}{\partial w_j}(J_{reg}^{\beta}(w_j))$$ # Assume a: learning rate.

$$w_j \leftarrow w_j - a \cdot \left[\frac{1}{N}\sum_{i=1}^{N}(y^i - t^i) \cdot x_j^i + \beta_j w_j\right]$$ # from ①

$$w_j \leftarrow w_j - \frac{a}{N}\sum_{i=1}^{N}(y^i - t^i) \cdot x_j^i - a\beta_j w_j$$

or $$w_j \leftarrow w_j - \frac{a}{N}\sum_{i=1}^{N}\left(\sum_{k=1}^{D} w_k x_k^i + b\right)x_j^i - a\beta_j w_j$$

② And the update rule for b is:

$$b \leftarrow b - a \cdot \frac{\partial}{\partial b}(J_{reg}^{\beta}(w))$$

$$b \leftarrow b - a \cdot \frac{\partial}{\partial b}\left(\frac{1}{2N}\sum_{i=1}^{N}(y^i - t^i)^2 + \frac{1}{2}\sum_{j=1}^{D}\beta_j w_j^2\right)$$

$$b \leftarrow b - a \cdot \left[\frac{1}{N}\sum_{i=1}^{N}(y^i - t^i) + 0\right]$$

because $\frac{\partial(y^i - t^i)}{\partial b} = (f(w) + b)' = 1$

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

Or $$b \leftarrow b - \frac{a}{N}\left[\sum_{k=1}^{D} w_k x_k^i + b\right]$$

Q3b

Q3b) Claim: I'm using k to replace j' in the handout

From Eq ①:

$$\frac{\partial \mathcal{J}^{\beta}_{reg}}{\partial w_j} = \frac{1}{N}\sum_{i=1}^{N}(y^i - t^i)\cdot x_j^i + \beta_j w_j$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left[\left(\sum_{k=1}^{D} w_k x_k^i\right) - t^i\right] x_j^i + \beta_j w_j$$

$$= \frac{1}{N}\sum_{k=1}^{D}\left(\sum_{i=1}^{N} x_j^i x_k^i\right) w_k - \frac{1}{N}\sum_{i=1}^{N} x_j^i t^i + \beta_j w_j$$

$$= \sum_{k=1}^{D}\left(\frac{1}{N}\sum_{i=1}^{N} x_j^i x_k^i\right) w_k + \beta_j w_j - \frac{1}{N}\sum_{i=1}^{N} x_j^i t^i$$

$$= \sum_{k=1}^{D-\{j\}}\left(\frac{1}{N}\sum_{i=1}^{N} x_j^i x_k^i\right) w_k + \left[\left(\frac{1}{N}\sum_{i=1}^{N} x_j^i x_j^i\right) + \beta_j\right] w_j - \frac{1}{N}\sum_{i=1}^{N} x_j^i t^i$$

Therefore,
$$A_{jk} = \begin{cases} \frac{1}{N}\sum_{i=1}^{N} x_j^i x_k^i, & \text{if } k \neq j \\ \left(\frac{1}{N}\sum_{i=1}^{N} x_j^i x_k^i\right) + \beta_j, & \text{if } k = j \end{cases}$$

$$c_j = \frac{1}{N}\sum_{i=1}^{N} x_j^i t^i$$

Q3c

Q3c. From b, $A_{jk} = \begin{cases} \frac{1}{N}\sum_{i=1}^{N} x_j^i x_k^i & , \text{if } k \neq j \\ (\frac{1}{N}\sum_{i=1}^{N} x_j^i x_k^i) + \beta_j & , \text{if } k = j \end{cases}$

$$C_j = \frac{1}{N}\sum_{i=1}^{N} x_j^i t^i$$

So, we can buid A as follows:

$$A = \begin{bmatrix} \frac{1}{N}\sum_{i=1}^{N} x_1^i x_1^i + \beta_1 & \frac{1}{N}\sum_{i=1}^{N} x_1^i x_2^i & \cdots & \frac{1}{N}\sum_{i=1}^{N} x_1^i x_D^i \\ \vdots & N\sum_{i=1}^{N} x_2^i x_2^i + \beta_2 & & \vdots \\ \vdots & & \ddots & \vdots \\ \frac{1}{N}\sum_{i=1}^{N} x_D^i x_1^i & \cdots & \cdots & \frac{1}{N}\sum_{i=1}^{N} x_D^i x_D^i + \beta_D \end{bmatrix}$$

$$C = \left[\frac{1}{N}\sum_{i=1}^{N} x_1^i t^i,\ \frac{1}{N}\sum_{i=1}^{N} x_2^i t^i,\ \dots,\ \frac{1}{N}\sum_{i=1}^{N} x_D^i t^i\right]^{-1}$$

Notice that, $\sum_{j'=1}^{D} A_{jj'} w_{j'}$ → it is multiplication of j' row in A with w.

So, $\sum_{j'=1}^{D} A_{jj'} w_{j'} - C_j = 0$

$\Rightarrow\ Aw - C = 0$ # each row

$\Rightarrow\ Aw = C$

$w = A^{-1} \cdot C$