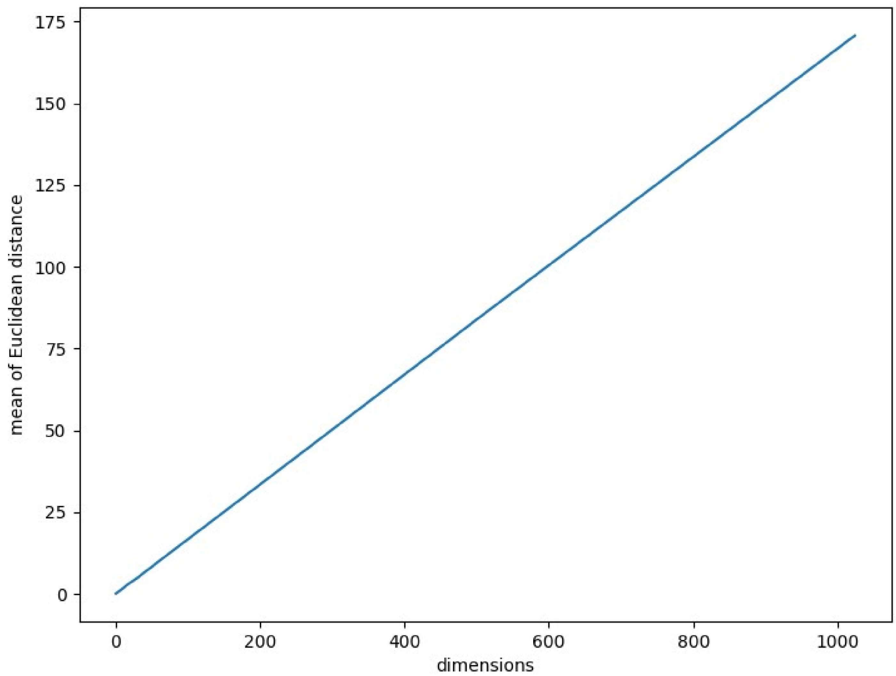
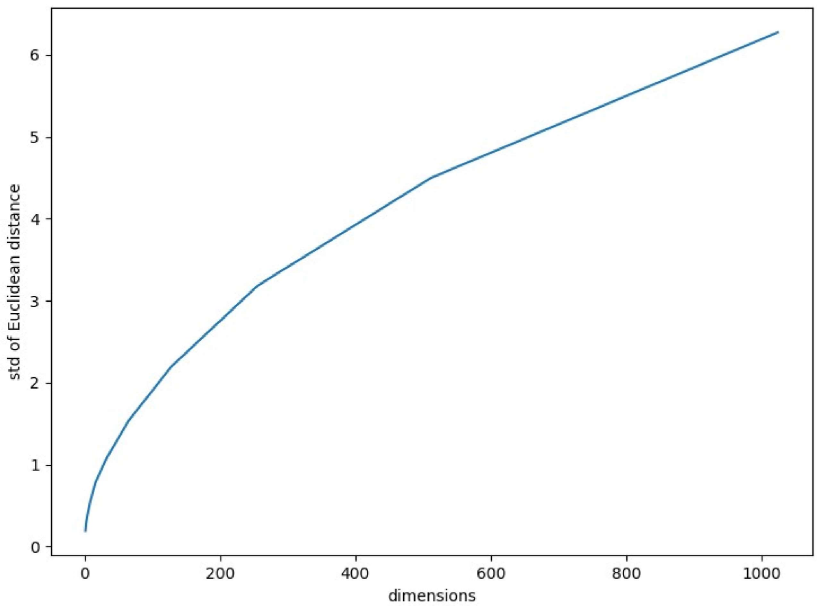
Q1 a> output for mean-dimension:



output for std-dimension:



Q1 b> assume all z_i are independent of each other;

Thus $E(R) = E\left(\sum_{i=1}^{d} z_i\right) = \sum_{i=1}^{d} E(z_i)$

$$= \sum_{i=1}^{d} \frac{1}{6} = \frac{d}{6}$$

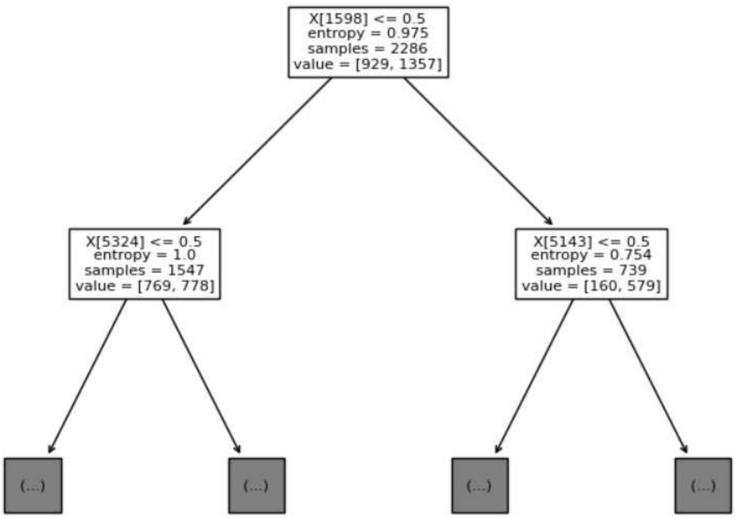
$$V(R) = V\left(\sum_{i=1}^{d} z_i\right) = \sum_{i=1}^{d} V(z_i) = \sum_{i=1}^{d} \frac{7}{180} = \frac{7d}{180}$$

Q2 b>

```
model: entropy; depth: 5; accuracy: 0.70816
model: gini; depth: 5; accuracy: 0.71224
model: entropy; depth: 12; accuracy: 0.72245
model: gini; depth: 12; accuracy: 0.76531
model: entropy; depth: 20; accuracy: 0.76531
model: gini; depth: 20; accuracy: 0.77143
model: entropy; depth: 30; accuracy: 0.78163
model: gini; depth: 30; accuracy: 0.76939
model: entropy; depth: 45; accuracy: 0.76735
model: gini; depth: 45; accuracy: 0.76122
```

← classifier of entropy with depth 30 yields highest accuracy;

c>



Q2d

```
keyword 1598: 0.05614365126920684
keyword 5143: 0.05278618037184435
keyword 5324: 0.04315750865439716
```

Q3 a>

$$\frac{\partial J_{reg}}{\partial w_j} = \frac{\partial}{\partial w_j}\left(\frac{1}{2N}\sum_{i=1}^{N}\left(\sum_{j'=1}^{D} w_{j'}x_{ij'}+b-t_i\right)^2+\frac{1}{2}\sum_{j'=1}^{D}\beta_{j'}w\right.$$

$$= \frac{1}{N}\sum_{i=1}^{N}\left(\sum_{j'=1}^{D} w_{j'}x_{ij'}+b-t_i\right)\cdot\frac{\partial}{\partial w_j}\left(\sum_{j'=1}^{D} w_{j'}x_{ij'}+b-t\right.$$

$$+\frac{\partial}{\partial w_j}\frac{1}{2}\beta_1 w_1^2+\cdots\frac{\partial}{\partial w_j}\frac{1}{2}w_j^2\beta_j+\cdots\frac{\partial}{\partial w_D}\frac{1}{2}w_D^2\beta_D$$

$$=\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{j'=1}^{D} w_{j'}x_{ij'}+b-t_i\right)x_{ij}+\beta_j w_j$$

thus $w_j \leftarrow w_j - \frac{\alpha}{N}\sum_{i=1}^{N}(y_i-t_i)x_{ij}-\beta_j w_j$

$$\frac{\partial J_{reg}}{\partial b}=\frac{\partial}{\partial b}\left(\frac{1}{2N}\sum_{i=1}^{N}\left(\sum_{j'=1}^{D} w_{j'}x_{ij'}+b-t_i\right)^2+\frac{1}{2}\sum_{j'=1}^{D}\beta_{j'}w_{j'}^2\right)$$

$$=\frac{1}{N}\sum_{i=1}^{N}(y_i-t_i)\cdot\frac{\partial}{\partial b}\left(\sum_{j'=1}^{D} w_{j'}x_{ij'}+b-t_i\right)$$

$$=\frac{1}{N}\sum_{i=1}^{N}(y_i-t_i)\cdot\left(\sum_{j'=1}^{D}0+1-0\right)=\frac{1}{N}\sum_{i=1}^{N}(y_i-t_i)$$

$$b\leftarrow b-\frac{\alpha}{N}\left(\sum_{i=1}^{N}(y_i-t_i)\right)$$

Q3 b>

$$\frac{\partial J_{reg}}{\partial w_j}=\frac{1}{2N}\left[\sum_{i=1}^{N}\left(\sum_{j'=1}^{d}w_{j'}x_{ij'}-t_i\right)^2\right]'+\frac{1}{2}\left[\sum_{j'=1}^{D}(\beta_j w_{j'}^2)\right]'$$

$$=\frac{1}{N}\sum_{i=1}^{N}\left(\sum_{j'=1}^{d}w_{j'}x_{ij'}-t_i\right)x_{ij}+\beta_j w_j$$

$$=\frac{1}{N}\left[\sum_{i=1}^{N}\left(w_1x_{i1}x_{ij}+\cdots w_jx_{ij}^2+\cdots w_dx_{id}x_{ij}-t_ix_{ij}\right)\right]+\beta_j w_j$$

$$= \frac{1}{N}\left[\sum_{j'=1}^{D} w_{j'}\left(\sum_{i=1}^{N} x_{ij'} x_{ij}\right) - \sum_{i=1}^{N} t_i x_{ij}\right] + \beta_j w_j$$

when $j' = j$ $\quad A_{jj} = \frac{1}{N}\sum_{i=1}^{N} x_{ij} x_{ij} + \beta_j$ $\qquad C_j = \frac{1}{N}\sum_{i=1}^{N} t_i x_{ij}$

$j' \neq j$ $\quad A_{jj'} = \frac{1}{N}\sum_{i=1}^{N} x_{ij'} x_{ij}$

Q3C> let $\vec{\beta} = [\beta_1, \cdots \beta_D]$;

by treating each entry of matrix A as the dot product of $\vec{x}$ s, (data vectors),

$$A = \frac{1}{N}\left(\begin{bmatrix} \sum_{i=1}^{N} x_{i1} x_{i1} & \cdots & \sum_{i=1}^{N} x_{iD} x_{i1} \\ \vdots & \ddots & \\ \sum_{i=1}^{N} x_{i1} x_{iD} & & \sum_{i=1}^{N} x_{iD} x_{iD} \end{bmatrix} + \vec{\beta} \cdot N \cdot I\right)$$

since $X = \begin{bmatrix} \vec{x}_1 \\ \vdots \\ \vec{x}_n \end{bmatrix}_{N \times D}$, then the dot product matrix is formed

by : $X^T X$; since X^T has dimension $D \times N$, thus

$X^T X$ has dimension $D \times D \Rightarrow$ can multiply with $\vec{w}$ $(1 \times D)$

Then $A = \frac{1}{N}(X^T X + \vec{\beta} N I)$;

$$C = \begin{bmatrix} C_1 \\ \vdots \\ C_D \end{bmatrix} = \frac{1}{N}\begin{bmatrix} \sum_{i=1}^{n} x_{i1} \cdot t_i \\ \vdots \\ \sum_{i=1}^{n} x_{iD} \cdot t_i \end{bmatrix} = \frac{1}{N}(X^T \cdot \vec{t}\,)$$

since X^T has dimension $D \times N$, which is multiplicable with $\vec{t}$ having dimension '$N \times 1$'

Thus $Aw = C \Rightarrow \quad w = A^{-1} C = (X^T X + \vec{\beta} N I)^{-1} \cdot (X^T \cdot \vec{t}\,)$