1.(a) $E(Z) = E(X^2+Y^2-2XY) = E(X^2)+E(Y^2)-2E(XY)$

Since $X \perp Y$ $= Var(X)+E(X)^2+Var(Y)+E(Y)^2-2E(X)E(Y)$

$= \frac{1}{12}+\frac{1}{4}+\frac{1}{12}+\frac{1}{4}-2\times\frac{1}{2}\times\frac{1}{2}$

$= \frac{1}{6}.$

$X, Y \sim U[0,1]$, $E(X) = \int_0^1 x = \frac{1}{2}$

$E(X^2) = \int_0^1 x^2 = \frac{1}{3}$

$E(X^3) = \int_0^1 x^3 = \frac{1}{4}$

$E(X^4) = \int_0^1 x^4 = \frac{1}{5}$

$Var(Z) = Var[(X-Y)^2] = E[(X-Y)^4] - E[(X-Y)^2]^2$

$= E(X^4) - 4E(X^3)E(Y) + 6E(X^2)E(Y^2) - 4E(X)E(Y^3) + E(Y^4) - E[(X-Y)^2]^2$

$= \frac{1}{5} - 4\times\frac{1}{4}\times\frac{1}{2} + 6\times\frac{1}{3}\times\frac{1}{3} - 4\times\frac{1}{2}\times\frac{1}{4} + \frac{1}{5} - \frac{1}{36}$

$= \frac{1}{15} - \frac{1}{36}$

$= \frac{7}{180}$

Hilroy

1. (b) $E(R) = E(Z_1 + \cdots + Z_d) = E(Z_1) + \cdots + E(Z_d)$

$= E((X_1 - Y_1)^2) + \cdots + E((X_d - Y_d)^2)$

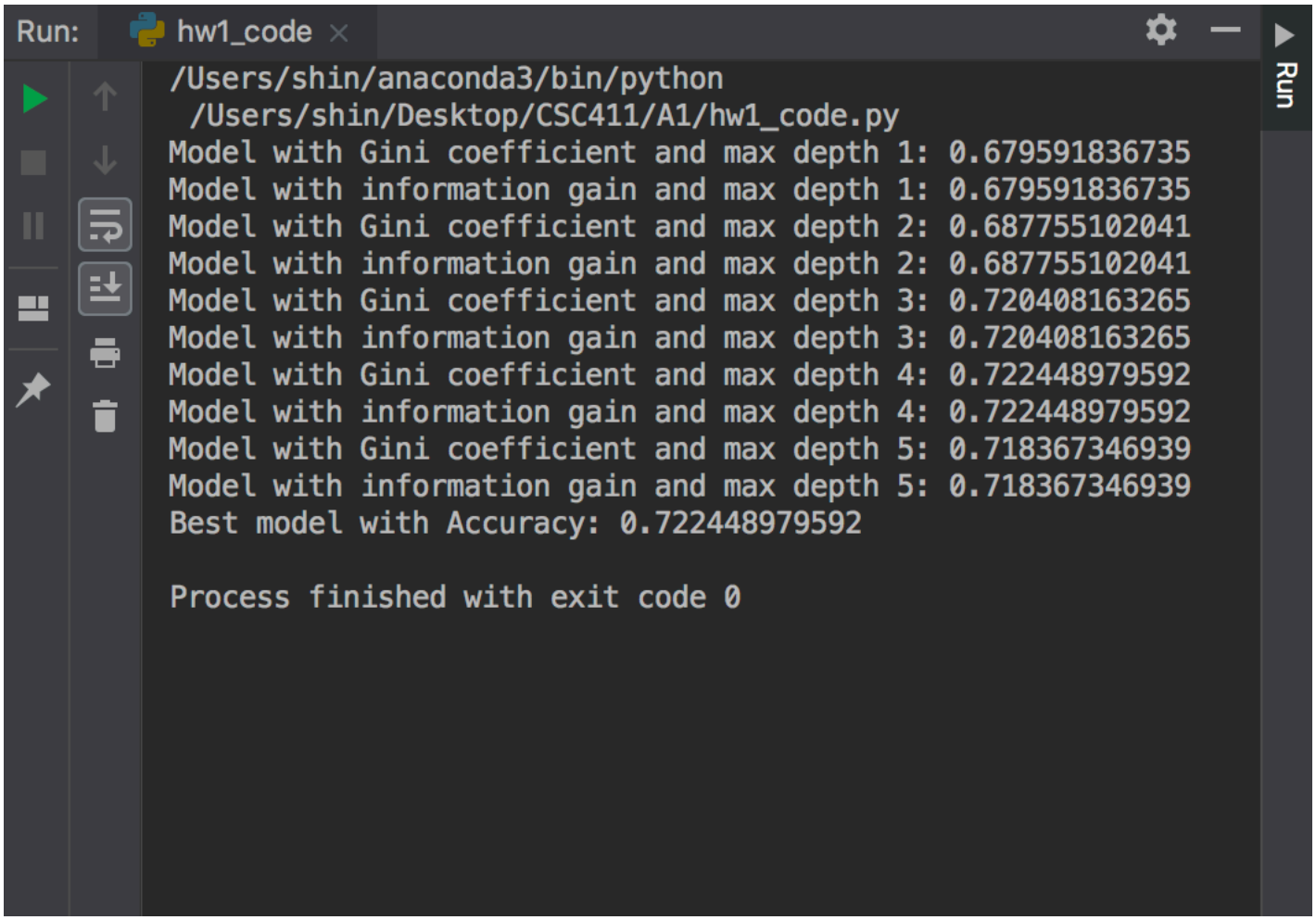
$= \frac{d}{6}$

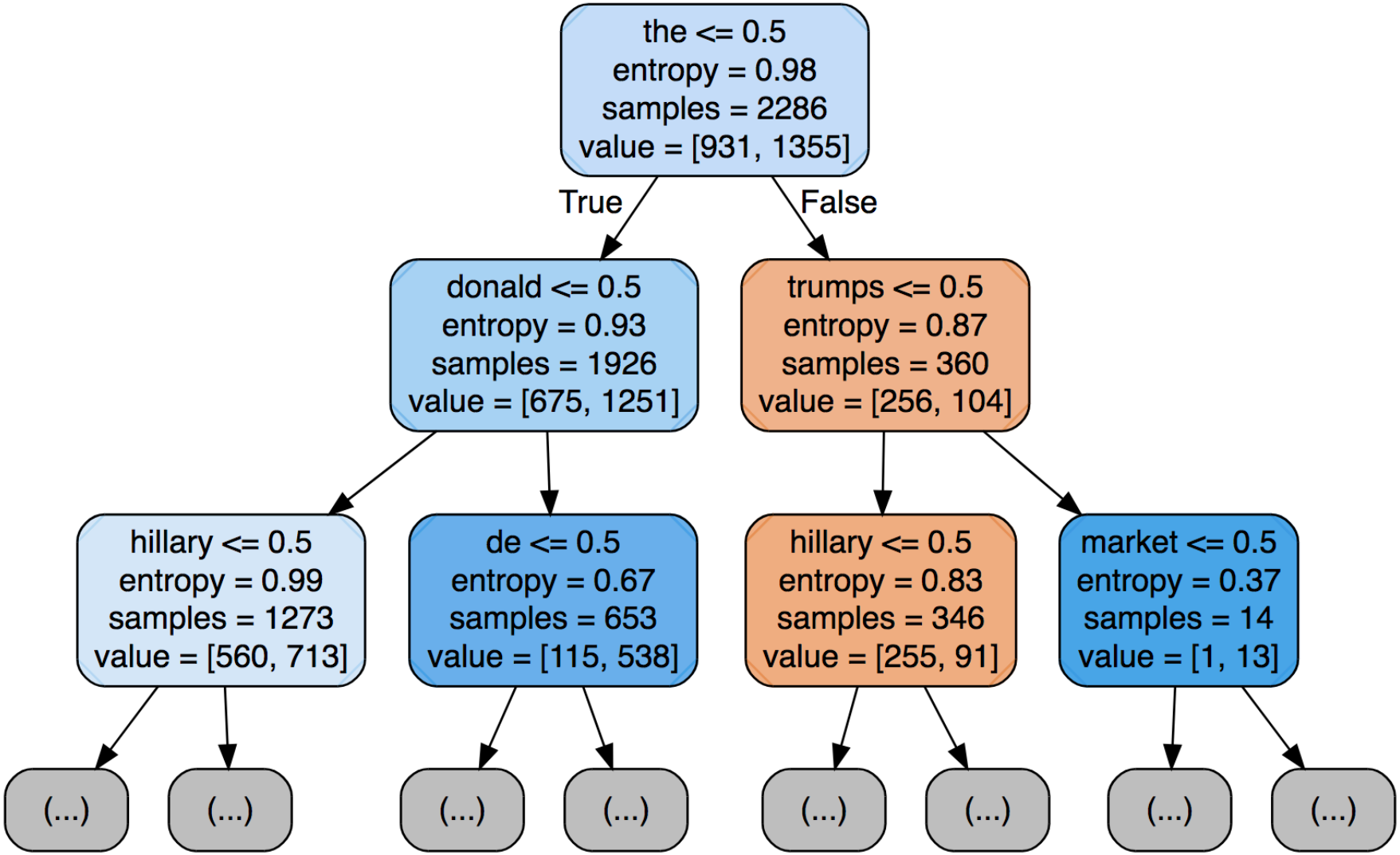
$Var(R) = Var(Z_1 + \cdots + Z_d) = Var(Z_1) + \cdots + Var(Z_d)$

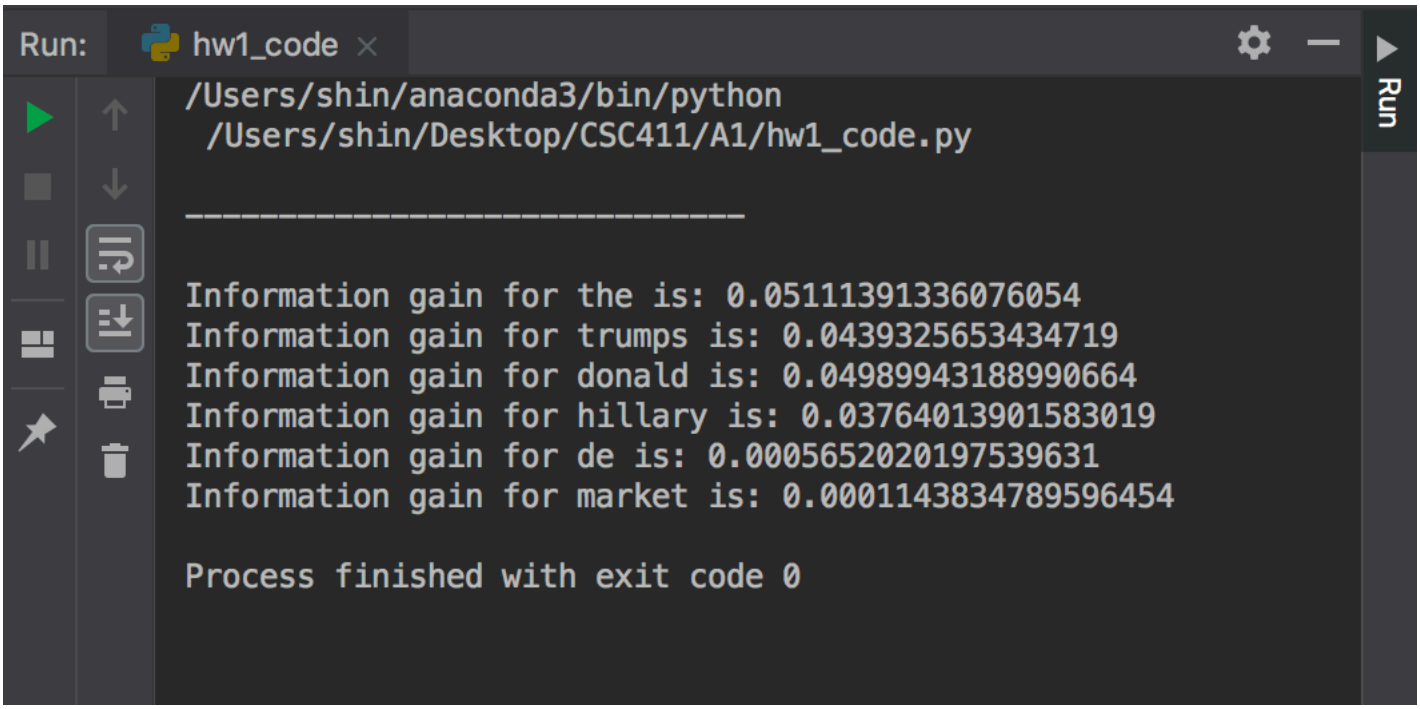
Since Z_i are iid.

$= V((X_1 - Y_1)^2) + \cdots + Var((X_d - Y_d)^2)$

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


Run:
hw1_code
Run
/Users/shin/anaconda3/bin/python
/Users/shin/Desktop/CSC411/A1/hw1_code.py
Model with Gini coefficient and max depth 1: 0.679591836735
Model with information gain and max depth 1: 0.679591836735
Model with Gini coefficient and max depth 2: 0.687755102041
Model with information gain and max depth 2: 0.687755102041
Model with Gini coefficient and max depth 3: 0.720408163265
Model with information gain and max depth 3: 0.720408163265
Model with Gini coefficient and max depth 4: 0.722448979592
Model with information gain and max depth 4: 0.722448979592
Model with Gini coefficient and max depth 5: 0.718367346939
Model with information gain and max depth 5: 0.718367346939
Best model with Accuracy: 0.722448979592
Process finished with exit code 0


the <= 0.5
entropy = 0.98
samples = 2286
value = [931, 1355]
True
False
donald <= 0.5
entropy = 0.93
samples = 1926
value = [675, 1251]
trumps <= 0.5
entropy = 0.87
samples = 360
value = [256, 104]
hillary <= 0.5
entropy = 0.99
samples = 1273
value = [560, 713]
de <= 0.5
entropy = 0.67
samples = 653
value = [115, 538]
hillary <= 0.5
entropy = 0.83
samples = 346
value = [255, 91]
market <= 0.5
entropy = 0.37
samples = 14
value = [1, 13]
(...)
(...)
(...)
(...)
(...)
(...)
(...)
(...)


Run:
hw1_code
/Users/shin/anaconda3/bin/python
/Users/shin/Desktop/CSC411/A1/hw1_code.py
Information gain for the is: 0.051113913360760 54
Information gain for trumps is: 0.0439325653434719
Information gain for donald is: 0.04989943188990664
Information gain for hillary is: 0.03764013901583019
Information gain for de is: 0.0005652020197539631
Information gain for market is: 0.0001143834789596454
Process finished with exit code 0
Run