Machine Learning

Tutorial 2

**Arneev Singh (2180393)**

**Shravan Singh (2173638)**

**Kiara Gabriel (2161334)**

**Phola Bavuma (1848739)**

Question 1

**a)**

ŷ = θ0+θ1

= 0.4 + 0.8(1)

ŷ(1) = 1.2  
  
*E*(θ) = (y- ŷ)2

=(1-1.2)2

*E*(θ) =

**b)**

ŷ = θ0+θ1

= 0.4 + 0.8(2)

ŷ(2) = 2  
  
*E*(θ) = (y- ŷ)2

= (3-2)2

*E*(θ) = 1

**c)**

ŷ = θ0+θ1

= 0.4 + 0.8(3)

ŷ(3) = 2.8  
  
*E*(θ) = (y- ŷ)2

= (2-2.8)2

*E*(θ) =

**d)**

ŷ = θ0+θ1

= 0.4 + 0.8(3)

ŷ(3) = 2.8  
  
*E*(θ) = (y- ŷ)2

= (2-2.8)2

*E*(θ) =   
  
ŷ = θ0+θ1

= 0.4 + 0.8(4)

ŷ(4) = 3.6  
  
*E*(θ) = (y- ŷ)2

= (3-3.6)2

*E*(θ) =   
  
ŷ = θ0+θ1

= 0.4 + 0.8(5)

ŷ(5) = 4.4  
  
*E*(θ) = (y- ŷ)2

= (5-4.8)2

*E*(θ) =   
  
  
*E*(θ) = ½ ( + 1 + + + ) = 1.04

**d)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 |
| ŷ = | 1 | 2 | 3 | 4 | 5 |
| y | 1 | 3 | 2 | 3 | 5 |

*E*(θ) = ½

*E*(θ) = 1.5

**e)**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | 1 | 2 | 3 | 4 | 5 |
| ŷ = 1+2 | 3 | 5 | 7 | 9 | 11 |
| y | 1 | 3 | 2 | 3 | 5 |

*E*(θ) = ½

*E*(θ) = 52.5

**f)**

Chart

Description automatically generated

ŷ = 0.4+0.8

*E*(θ) = 1.04

Chart, scatter chart

Description automatically generated

ŷ =

*E*(θ) = 1.5

Chart

Description automatically generated

ŷ = 1+2

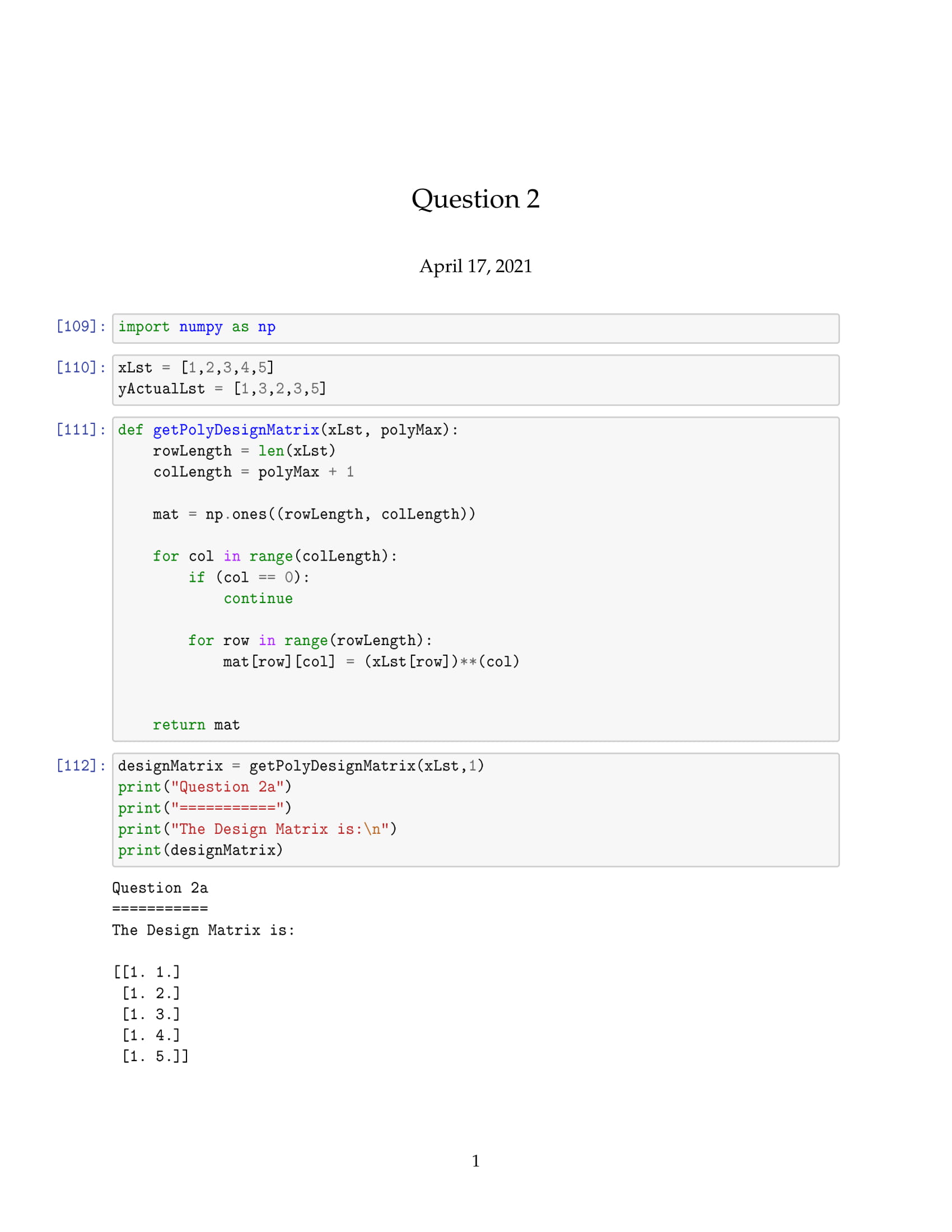
*E*(θ) = 52.5

**g)**

The best model is ŷ = 0.4+0.8

ŷ(6) = 0.4 + 0.8(6)

ŷ(6) = 5.2





Question 3

H(p) = -(p1 log2(p1) + p2 log2(p2))

= -(0.5 log2(0.5) + 0.5 log2(0.5))

= 1

H(p) = -(p1 log2(p1) + p2 log2(p2))

= -( log2() + log2())

= 0.918

H(p) = -(p1 log2(p1) + p2 log2(p2))

= -(0.25 log2(0.25) + 0.75 log2(0.75))

= 0.811

H(p) = -(p1 log2(p1) + p2 log2(p2))

= -(0 log2(0) + 1 log2(1))

= 0

H(p) = -(p1 log2(p1) + p2 log2(p2) + p3 log2(p3))

= -( log2() + log2() + log2())

= 1.584

H(p) = -(p1 log2(p1) + p2 log2(p2) + p3 log2(p3))

= -(0.25 log2(0.25) + log2(0.25) + log2())

= 1.5

H(p) = -(pY log2(pY) + pN log2(pN))

= -( log2( ) + log2( ))

= 1

H(p) = -(pY log2(pY) + pN log2(pN))

= -( log2( ) + log2())

= 0.543

Question 4

1. H(D) = -(pP log2(pP) + pF log2(pF))

= -( log2() + log2())

= *0.9544*

1. F1(COMS2 mark)

* For F1 = A

PP = ; PF =

H(DA) = -( log2() + log2())

= **1**

* For F2 = B

PP = ; PF =

H(DB) = -( log2() + log2())

= **0.9182**

* For F2 = C

PP = ; PF =

H(DC) = -( log2() + log2())

= **0.9182**

* *Gain(D, F1)*

= H(D) -

= 0.9544 - ((2 1) + (3 0.918) + (3 0.918)

= **0.0159**

1. F2(doing labs)

* For F2 = Y

PP = ; PF =

H(DY) = -( log2() + log2())

= **0.811**

* For F2 = N

PP = ; PF =

H(DN) = -( log2() + log2())

= **1**

* *Gain(D, F2)*

= H(D) -

= 0.9544 - ((4 0.811) + (4 1))

= **0.0489**

1. F3(doing tuts)

* For F3 = Y

PP = 1 ; PF = 0

H(DY) = -( log2() + log2())

= **0**

* For F2 = N

PP = ; PF =

H(DN) = -( log2() + log2())

= 0.9709

* *Gain(D, F2)* = H(D) -

= 0.9544 - ((3 0) + (5 0.9709))

= **0.3475**

1. The variable with the maximum gain is ***doing tuts***because **0.3475**>0.0489>0.0159. The node will have two branches {Y, N}.
2. We repeat the process to maximize the prediction. As we extend the tree into different branches the data point are also separated, leaving us with smaller data samples that need their own respective sub-tree. So, for each branch they will each be calculated on a new set of sample data.

Since we used a feature as the root node, there will be one less feature in the dataset. We treat each branch separate since there are different decisions/features that are not linked to neighbouring decisions/features. Basically, sibling roots do not affect one another.