Kevin Ngo
307385434

Homework 5.

Question 2:

a) We have target: $\theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3 + \theta_4 X_1 X_2 + \theta_5 X_1 X_3$

with: X_1 = GPA; X_2 = Age; X_3 = Type of position $\begin{cases} 1: \text{Technical} \\ 0: \text{Non-Technical} \end{cases}$

→ To compare the salary between Technical and Non-Technical position, we only need to compare $\theta_3 X_3 + \theta_5 X_1 X_3$ (fixed GPA and Age)

→ We have $\theta_3 = -30$ and $\theta_5 = 10 \Rightarrow -30X_3 + 10X_1X_3$.

* For Non-Technical position $\Rightarrow X_3 = 0 \Rightarrow -30X_3 + 10X_1X_3 = 0$ (salary)
* For Technical position $\Rightarrow X_3 = 1 \Rightarrow -30 + 10X_1$ (salary)

↳ So, the salary of Technical is higher or lower than Non-Technical (0) depends on X_1 which is GPA.

-> We can see that ~~GPA < 3~~ we have 3 cases:

1. if GPA = 3 ⇒ Technical and Non-Technical earn the same (0)
2. if GPA < 3 ⇒ Technical earns less than Non-Technical
3. if GPA > 3 ⇒ Technical earns more than Non-Technical

⇒ The answer iii is correct ⇒ For a fixed value of Age and GPA, Technical positions earn more on average than Non-Technical ~~pos~~ positions when GPA is high enough (GPA > 3)

b) Predition the salary of a Technical and NonTechnical positions with Age = 25, GPA = 4.0

⇒ Technical: $\theta_0 + \theta_1 X_1 + \theta_2 X_2 + \theta_3 X_3 + \theta_4 X_1 X_2 + \theta_5 X_1 X_3 = 40 + 20\times4 + 0.1\times25 + (-30)\times1 + 0.01\times25\times4 + 10\times4\times1 = 133.5$

Non-Technical: $40 + 20\times4 + 0.1\times25 - 30\times0 + 0.01\times4\times25 + 10\times4\times0 = 123.5$.

Question 3: Matrix-to-Vector ~~using~~ multiplication using MapReduce with 4 mappers and 2 reducers

$$\begin{bmatrix} 5 & -3 & 3 & 7 \\ 4 & 2 & -8 & 6 \end{bmatrix} \begin{bmatrix} 3 \\ 7 \\ -9 \\ 2 \end{bmatrix}.$$

→ 4 mappers ⇒ correspond to 4 cols ⇒ mapper j processes row i of the matrix at a time.

→ input : (i, j) , m_{ij} 4

– output : i, $m_{ij} \times v_j$

i: row

j: col (mapper)

	Input		Output	
Mappers.	In-Key (i,j)	In-value (m_{ij})	out-key (i)	intermediate values ($m_{ij} \times v_j$)
Mapper 1 ($j=1$)	→ $i=1$, $j=1$ → $(1,1)$	5	1	15
	→ $i=2$, $j=1$ → $(2,1)$	4	2	12
Mapper2 ($j=2$)	→ $i=1$, $j=2$ → $(1,2)$	−3	1	−21
	→ $i=2$, $j=2$ → $(2,2)$	2	2	14
Mapper 3 ($j=3$)	→ $i=1$, $j=3$ → $(1,3)$	3	1	−27
	→ $i=2$, $j=3$ → $(2,3)$	−8	2	72
Mapper 4 ($j=4$)	→ $i=1$, $j=4$ ⇒ $(1,4)$	7	1	14
	→ $i=2$, $j=4$ → $(2,4)$	6	2	12

→ 2 reducers ⇒ reducer i receives $(\underset{\text{out\_key}}{i}, \overbrace{[m_{i1}v_1, \dots, m_{iN}v_N]}^{\text{List of all intermediate values from mappers (all)}})$

→ sum all values → $R_i = \sum_j m_{ij} \times v_j$ = out_value.

	Input		Output	
Reducers	out_key (i)	list of all $[m_{i1}v_1, m_{i2}v_2, \dots, m_{iN}v_N]$	out_key (i)	out_value (sum of all element in list)
Reducer 1 ($i=1$)	1	[15, −21, −27, 14]	1	−19
Reducer2 ($i=2$)	2	[12, 14, 72, 12]	2	110

$$\Rightarrow R = \begin{bmatrix} -19 \\ 110 \end{bmatrix} \text{ with } R_1 = -19 \ ; \ R_2 = 110$$