**AIAP Question 3**

A. **Gradient Descent**

1. $h\theta(x) = \theta 0 + \theta 1x1 + \theta 2x2$
2. MSE Cost function: $1/n \sum_{i=1}^{n}(h\theta(x^{(i)}) - y^{(i)})^2$
3. We have 3 simultaneous equations:

$30 = \theta 0 + 3\theta 1 + \theta 2$ ——— (1)

$55 = \theta 0 + 6\theta 1 + \theta 2$ ——— (2)

$70 = \theta 0 + 3\theta 1 + 3\theta 2$ ——— (3)

(3) - (2) = $3\theta 1 = 25$

Thus $\theta 1 = 25/3$ —— (4)

Sub (4) into (3), we have $45 = 3\theta 2 + \theta 0$ —— (5)

(1)X 3 - (5), we get $\theta 0 = -15$

Thus $\theta 2 = 10$

**B. Regularization**

1. Regularization refers to the process where we try to prevent overfitting the model by reducing the coefficient values towards 0. This fitting procedure includes a loss function that is known as residual sum of squares. Coefficients are chosen to minimize the loss function. Unregularized cost functions may not generalize well with future data. For example, a simple regression formula is $y = a1x1 + a2x2 + b$ In this case if coefficient a1 is of a much higher value than a2, then variations in a2 may not be reflected accurately as the predictions will rely on changes in a1. $\lambda$ as a parameter with which we decide on the how much we want to penalize the flexibility of the model. With $\lambda$ = 0, there is no penalty and this gives a least squares estimate. But as $\lambda \rightarrow \infty$ , the penalty grows and the coefficient values approach 0

2. L1 and L2 regularisation refer to Lasso and Ridge regression respectively. The difference is that Lasso penalizes only the high coefficients. Lasso can eliminate all the coefficients in comparison to Ridge.