HW1

Problem 1 i) The equation we have is: $Y = X\beta + \varepsilon$

Where:

- Y is an $n \times 1$ matrix
- X is an $n \times (k+1)$ matrix)
- β is a $(k+1) \times 1$ matrix
- ε is an $n \times 1$ matrix

Rewritting the equation to see the dimensions:

$$Y_{n\times 1} = X_{n\times(k+1)} \cdot \beta_{(k+1)\times 1} + \varepsilon_{n \times 1}$$

The multiplication of vectors X and β gives us an $n \times 1$ vector. This is because the number of columns in X $(k+1)$ is the same as the number of rows in β $(k+1)$. The product of these vectors results in an $n\times 1$ matrix which can be added to ε (also an $n \times 1$ matrix) which matches the dimensions of Y (also an $n \times 1$ matrix).

ii) We want to write the sum of squared residuals (MSE) in matrix form given that the residuals are given by $e = Y - X\beta$ and the sum of squared residuals is $e^T e$.

$$MSE = e^T e = (Y - X\beta)^T (Y - X\beta)$$

To make the next step easier, we will now expand the expression:

$$MSE = (Y - X\beta)^T (Y - X\beta)$$
$$= (Y^T - \beta^T X^T)(Y - X\beta)$$
$$= Y^T Y - Y^T X\beta - \beta^T X^T Y + \beta^T X^T X\beta$$

iii) Differentiating the sum of squared residuals obtained in step 2 with respect to β:

$$\frac{\partial MSE}{\partial \beta} = -Y^T X - X^T Y + 2X^T X\beta^*$$
$$= -2X^T Y + 2X^T X\beta^*$$

Setting the above derivative equal to zero:

$$-2X^T Y + 2(X^T X)\beta^* = 0 \Rightarrow (X^T X)\beta^* = X^T Y$$

iv) Finally, we want to show that $\beta^* = (X^TX)^{-1}(X^TY)$.
From part (iii) we had:

$$(X^TX)\beta^* = X^TY$$

Multiplying both sides of the above equation by the inverse of (X^TX):

$$(X^TX)^{-1}(X^TX)\beta^* = (X^TX)^{-1}X^TY$$

The product of a matrix and its inverse is the identity matrix:

$$I\beta^* = (X^TX)^{-1}X^TY$$

which can be simplified as

$$\beta^* = (X^TX)^{-1}X^TY$$

Problem 2: Please see attached code file.
Parameter estimates from manual linear regression:
Intercept = 2.93889
TV = 0.0457646
Radio = 0.18853
Newspaper = -0.00103749

Parameter estimates from scikit-learn's linear regression:
Intercept = 2.93889
TV = 0.0457646
Radio = 0.18853
Newspaper = -0.00103749

Therefore, the parameter estimates obtained from the two procedures are the same.