HW 1 USC ID: 6648517090 Name: Yu-Wei Chen

1.

For height = 150, 3 nearest neighbor : 150, 168, 171

$$\hat{y} = \frac{65+78+80}{3} = \frac{223}{3} = \boxed{74.33\ kg}$$

For height = 155, 3 nearest neighbor : 150, 168, 171

$$\hat{y} = \frac{65+78+80}{3} = \boxed{74.33\ kg}$$

For height = 165, 3 nearest neighbor : 168, 171, 178

$$\hat{y} = \frac{78+80+83}{3} = \frac{241}{3} = \boxed{80.33\ kg}$$

For height = 190, 3 nearest neighbor = 178, 182, 191

$$\hat{y} = \frac{83+80+100}{3} = \frac{263}{3} = \boxed{87.67\ kg}$$

2.

For height = 150, 3 nearest neighbor = 150, 168, 171

$$W_1 = \frac{1}{150-150} = \infty$$

$$W_2 = \frac{1}{|150-168|} = \frac{1}{18} = 0.056$$

$$W_3 = \frac{1}{|150-171|} = \frac{1}{21} = 0.048$$

$$\boxed{\hat{y} = 65\ kg}$$

For height = 155, 3 nearest neighbor : 150, 168, 171

$$W_1 = \frac{1}{|155-150|} = \frac{1}{5} = 0.2$$

$$W_2 = \frac{1}{|155-168|} = \frac{1}{13} = 0.077 \qquad \hat{y} = \frac{0.2 \times 65 + 0.077 \times 78 + 0.063 \times 80}{0.2 + 0.077 + 0.063}$$

$$W_3 = \frac{1}{|155-171|} = \frac{1}{16} = 0.063 \qquad = \frac{24.046}{0.34} = \boxed{70.72\ kg}$$

For height = 165, 3 nearest neighbor : 168, 171, 178

$$W_1 = \frac{1}{|165-168|} = \frac{1}{3} = 0.33 \qquad \hat{y} = \frac{0.33 \times 78 + 0.16 \times 80 + 0.077 \times 83}{0.33 + 0.16 + 0.077}$$

$$W_2 = \frac{1}{|165-171|} = \frac{1}{6} = 0.16$$

$$W_3 = \frac{1}{|165-178|} = \frac{1}{13} = 0.077 \qquad = \frac{44.931}{0.567} = \boxed{79.24\ kg}$$

For height = 190, 3 nearest neighbor = 178, 182, 191

$$W_1 = \frac{1}{|190-178|} = \frac{1}{12} = 0.083 \qquad \hat{y} = \frac{0.083 \times 83 + 0.125 \times 80 + 1 \times 100}{0.083 + 0.125 + 1}$$

$$W_2 = \frac{1}{|190-182|} = \frac{1}{8} = 0.125 \qquad = \frac{116.889}{1.208} = \boxed{96.76\,kg}$$

$$W_3 = \frac{1}{|190-191|} = 1$$

3.

$$J(x) = x^T Q x + d^T x + c$$

$$\nabla_x J(x) = \frac{\partial J}{\partial x}$$

$$\nabla_x(x^T Q x) = 2Qx, \quad \nabla_x(d^T x) = d \qquad \nabla_x(c) = 0$$

$$\nabla_x J(x) = 2Qx + d$$

$$H = \frac{\partial^2 J}{\partial x \partial x^T} \qquad \nabla_x J(x) = 2Qx + d$$

$$\frac{\partial 2Qx+d}{\partial x} = 2Q \qquad H = 2Q$$

4.

In ordinary Least Squares regression, we estimate the coefficient vector β as:

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

$$\hat{y} = x'\beta \Rightarrow \hat{y} = x'(X^TX)^{-1}X^Ty$$

$$= \sum_{i=1}^{n} [x'(X^TX)^{-1}X^T]_i \, y_i$$

$[\;]_i$ represents the i-th row of the feature Matrix

$$X_{n\times(p+1)} = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{bmatrix} \qquad y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

prediction $\hat{y}$ is a weighted sum of the train label y_i, where the weights are determined by the term inside the brackets.

If we set $k=n$ in KNN regression, meaning we use all training points, and we allow for a specific weighting scheme, then the prediction formula starts resembling linear regression.

In this sense, linear regression can be seen as a global, weighted form of KNN regression, where all training points influence the prediction, rather than just the nearest neighbors.

5.

$\beta = (X^T X)^{-1} X^T y$

$\hat{y} = X\beta$ $\qquad$ $\beta \in R^{p+1} \rightarrow X\beta$ represents a linear combination of the column of X.

The term $X(X^T X)^{-1} X^T$ ensure that $\hat{y}$ is always in the column space of X. $\hat{y} \in Col(X)$

This confirms that the predicted values in linear regression are simply a linear combination of the columns of X, ensuring that $\hat{y}$ lies in the space spanned by X.

Thus, $\hat{y}$ is a member of the column space of X.

6.

$$\frac{\partial}{\partial \beta} RSS(\beta) = \frac{\partial}{\partial \beta}\left((y - X\beta)^T (y - X\beta)\right) = 0$$

$$= \frac{\partial}{\partial \beta}(y^T y - y^T X\beta - \beta^T X^T y + \beta^T X^T X\beta) = 0$$

$$= \frac{\partial}{\partial \beta}(y^T y - 2\beta^T X^T y + \beta^T X^T X\beta) = 0$$

$$= -2X^T y + 2X^T X\beta = 0$$

$$= -2X^T (y - X\beta) = 0$$

$\hat{\beta} = (X^T X)^{-1} X^T y$

$\hat{y} = X\hat{\beta} = X(X^T X)^{-1} X^T y$

Residual $= y - \hat{y} = y - X(X^T X)^{-1} X^T y$

$$X^T(y - X\hat{\beta}) = X^T(y - X(X^TX)^{-1}X^Ty)$$
$$= X^Ty - X^TX(X^TX)^{-1}X^Ty$$
$$= X^Ty - X^Ty = 0$$

This equation proves that the residuals are orthogonal to the column space of X, ensuring the linear regression minimizes the sum of squared error correctly.