Michael Kwan
Machine Learning
HW 3

1a.
i. non-linear
ii. no underfitting
iii. $w_0 = 1 \quad w_1 = 2 \quad w_2 = 0$

b.
i. non-linear
ii. underfitting
iii. n/a

c.
i. non-linear
ii. underfitting
iii. n/a

2a. Model 1.
x_1 = cancer volume
$\hat{y} = w_0 + w_1 x_1$

Model 2.
x_1 = cancer volume, x_2 = patient age
$\hat{y} = w_0 + w_1 x_1 + w_2 x_2$

Model 3.
x_1 = patient age x_2 = cancer volume x_3 = cancer type
$\hat{y} = w_0 + w_1 x_1 + w_2 x_2 x_3 + w_3 x_2 (1 - x_3)$

b. 1 parameter in model 1 and 2 parameters in Model 2
Model 2 is the most complex

c. Model 1 = $\begin{bmatrix} 1 & .7 \\ 1 & 1.3 \\ 1 & 1.6 \end{bmatrix}$ Model 2 = $\begin{bmatrix} 1 & .7 & 55 \\ 1 & 1.3 & 65 \\ 1 & 1.6 & 70 \end{bmatrix}$ Model 3 = $\begin{bmatrix} 1 & .7 & 55 & 1 \\ 1 & 1.3 & 65 & 0 \\ 1 & 1.6 & 70 & 0 \end{bmatrix}$

d. Model 2 is best because of the lowest test MSE.

3. Figures 1 and 2 are underfit and Figure 3 is overfit because the val error is significantly greater than the training error of Figure 3 when the set size is small. In all three models as the data set size increased, they approached the same error.

4a. A flexible statistical learning method is better because a larger N will mean there is less variance and it will not have to worry about too many features.

b. An inflexible statistical learning method is better because it is easier to calculate for lots of features when the model is simpler. Also, a flexible model needs a lot of data to be effective.

c. A flexible statistical learning method is better because they are expected to have high variance with low bias. If an inflexible model had high variance and high bias, it is objectively worse.

5. $w_{ridge} = (X^T X + N\lambda I)^{-1} X^T y$

$$X = \begin{bmatrix} 1 & 6.6 & 1 & 4 \\ 1 & 6.4 & 2 & 5 \\ 1 & 7.2 & 2 & 5 \\ 1 & 6.4 & 2 & 5 \\ 1 & 7.2 & 2 & 5 \end{bmatrix} \quad y = \begin{bmatrix} 24 \\ 21.6 \\ 34.7 \\ 21.6 \\ 34.7 \end{bmatrix} \quad N = 3 \quad \lambda = 0.1$$

6. $\nabla E_{ridge}(w) = \frac{2}{N}(X^T X w - X^T y) + 2\lambda I w$