Jack Lanchantin
CS 6501 HW2

1.1

$$
\begin{aligned}
J(\beta) &= (y - X\beta)^2 + \lambda\beta\beta \\
&= (y - X\beta)^T(y - X\beta) + \beta^T(\lambda I)\beta \\
&= y^T y - 2yX\beta + X^T X\beta^T\beta + \beta^T(\lambda I)\beta
\end{aligned}
$$

$$
\begin{aligned}
\nabla_\beta J(\beta) &= -2y^T X + 2X^T X\beta + 2\lambda I\beta \\
0 &= -2y^T X + \beta(2X^T X + 2\lambda I) \\
2y^T X &= \beta(2X^T X + 2\lambda I) \\
\beta &= (X^T X + \lambda I)^{-1} y^T X
\end{aligned}
$$

1.2

$$
X = \begin{bmatrix} 1 & 2 \\ 3 & 6 \\ 5 & 10 \end{bmatrix}, Y = \begin{bmatrix} 1 & 2 & 3 \end{bmatrix}^T
$$

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

$$
X^T X = \begin{bmatrix} 5 & 15 & 25 \\ 15 & 45 & 75 \\ 25 & 75 & 125 \end{bmatrix} \longrightarrow \textit{non invertable}
$$

=> Cannot be solved by linear regression because $X^T X$ is non-invertable

1.3 Lasso Regression should be used because it can generate a sparse β vector

1.5 Ridge regression performs better on this data set. This is probably due to the fact that after running linear regression between x_1 and x_2, I found that $\theta \approx 2$. This means that the features are not linearly independent ($x_2 \approx 2x_1$), and thus the $X^T X$ matrix is non invertable, so using λI in the ridge regression to regularize the function makes it so that the $X^T X$ matrix is guaranteed to be invertable, resulting in a more accurate β.