1. **[3 Points]** Write the stochastic gradient descent update rules for w_j and b for ridge regression.

$L \approx (\hat{y}^{(i)} - y^{(i)})^2 + \lambda \sum_{j=1}^{m} w_j^2$

$= ((\sum w_j x_j^{(i)} + b) - y^{(i)})^2 + \lambda \sum_{j=1}^{m} w_j^2$

$\frac{dL}{dw_j} = [2(\hat{y}^{(i)} - y^{(i)}) x_j + 2\lambda w_j]$

$\Rightarrow w_j \leftarrow w_j - \eta(2(\hat{y}^{(i)} - y^{(i)}) x_j + 2\lambda w_j)$

$\Rightarrow w_j \leftarrow (1 - 2\lambda\eta) w_j - 2\eta(\hat{y}^{(i)} - y^{(i)}) x_j$

$b \leftarrow b - \eta \times 2(\hat{y}^{(i)} - y^{(i)})$

5.4.2 Lasso regression

The loss for L_1 regularization can be written as

$$\mathcal{L} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}^{(i)} - y^{(i)})^2 + \lambda \sum_{j=1}^{m} \|w_j|$$

Note that we do not regularize b. With the SGD approximation, this becomes

$$\mathcal{L} \approx (\hat{y}^{(i)} - y^{(i)})^2 + \lambda \sum_{j=1}^{m} |w_j|$$

1. **[3 Points]** Write the stochastic gradient descent update rules for w_j and b for lasso regression.

if $w_j \geq 0$:

$w_j \leftarrow w_j - 2\eta(\hat{y}^{(i)} - y^{(i)}) x_j - \eta\lambda$

else:

$w_j \leftarrow w_j - 2\eta(\hat{y}^{(i)} - y^{(i)}) x_j + \eta\lambda$

$b \leftarrow b - 2\eta(\hat{y}^{(i)} - y^{(i)})$