Problem 1

a. LLS regression
    i. The design matrix X is 11 x 3
    ii. The parameter vector w is 3 x 1
    iii. The product X'X is 3 x 3
b. For 1000 point dataset
    i. The design matrix is 1000 x 3
    ii. The parameter vector w is 3 x 1
    iii. The size of the product X'X is 3 x 3
c. False: The model complexity is determined by the number of model parameters.
True: The number of data points used for training would help in increasing the model accuracy.

Problem 2

Since the function slope changes twice, a 3rd order polynomial ($w_1x_1^3 + w_2x_1^2 + w_3x_1 + w_4$) model would be able to describe these points quite well.

Problem 3

$$x_{new} = x_{old} + \eta \cdot \frac{\partial \text{obj}}{\partial x}$$

Since we want to reach the top of the mountain, we must follow the path that leads us up the slope of the mountain. Therefore, our next step should be in the direction of positive gradient/slope. Therefore, the formula attached above should lead us up the slope.

Problem 4

1. Increasing the batch size does not have any direct effect on the oscillations while minimizing the objective function.
2. Switching to stochastic gradient descent can help partially as it picks random samples and this might cause the gradient to smoothen out over time.
3. Yes, reducing the learning rate can help in reducing the oscillations as it prevents the update from overshooting.

Problem 5

On reducing the L1_ratio term to 0, the formulant becomes equivalent to an LLS with L2 regularization problem. The ($1 - L1_{ratio}$) coefficient becomes 1, and the associated norm is $||w||_2^2$, which is an L2 norm.