A) Based on plotting the training and test data, I can see that both data sets are fairly linear. This means that linear regression will be pretty effective in predicting the label y with the features of x. We can visualize a fairly linear relationship between the data already.

B)

```
if m==1:
    # ============================================================ #
    # YOUR CODE HERE:
    # IMPLEMENT THE MATRIX X_out with each entry = [1, x]
    X_out[:,0] = 1
    X_out[:,1] = X.reshape((n,))
    # ============================================================ #
    pass
    # ============================================================ #
```

C)

```
if m==1:
    # ============================================================ #
    # YOUR CODE HERE:
    # PREDICT THE TARGETS OF X
    X_new = self.get_poly_features(X)
    y_pred = np.dot(X_new, self.theta)
    # ============================================================ #
    pass
    # ============================================================ #
```

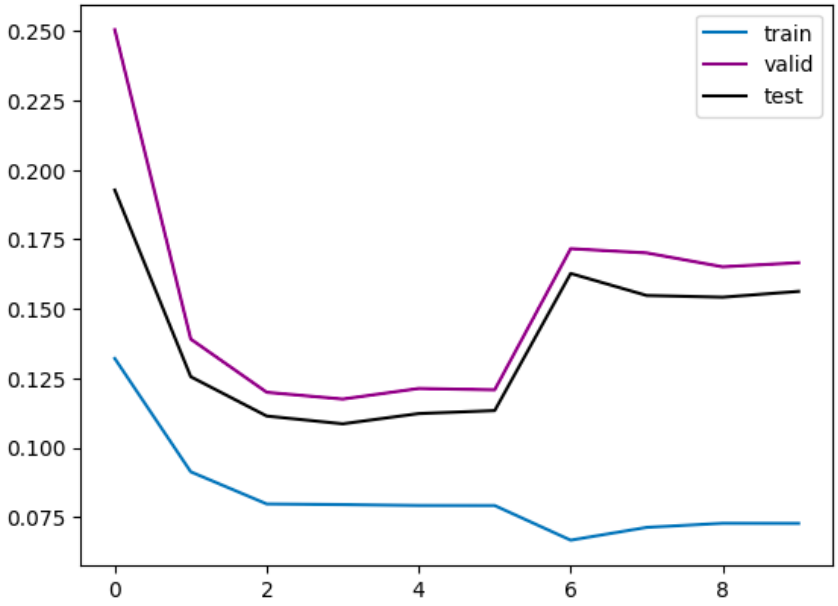
D) All learning rates lead to convergence eventually, although they vary in how fast they reach convergence.

E)

```
Optimal solution loss 0.13208969101982215
Optimal solution theta [[-0.37906992]
 [ 0.8852483 ]]
```

These results are fairly similar from our original MBGD calculations, which had a loss of 0.13458, which is only slightly bigger than the loss presented by the closed form solution (listen above). Similarly, the MBGD theta values are -0.2444 and 0.8397, which are somewhat similar to the ones found above. The 0.8.. term is quite similar, but the other theta value varied by 0.1, which is quite significant.

F) Here is the plot:



The polynomial that best fits the data is m = 3, as it has the lowest amount of loss for all three datasets. We can see a big gap in the training error and validation error for m = 6 and above. This indicates high variance, and thus overfitting.

G) We want the lambda that gives us the lowest test loss value. In this case, we would want the lambda value at index 7 (of 1 to 10), which is the lambda value of 10^(-3).