Problem 2.

1.

The least squares problem is a regression method to determine a solution for an overdetermined data set. Essentially, when determining a regression line, the goal is to minimize the squares of the residuals, so you end up with the “least squares”. What this does, is it weights outliers more, because large residuals get squared and are favored. Least squares is favored because of its relationship to the natural gaussian distribution. If errors are normal distributed among the regression line, then the least squares fit is considered optimal. So, least square fit can work well in many real-life scenarios because there are a lot of cases where the noise (error) is normal distributed.

2. http://mathworld.wolfram.com/LeastSquaresFitting.html

Problem 4.

1.

Regularization introduces new information to prevent overfitting. L1 and L2 norm are forms of regularization. Regularization reduces variance in a regression model, without losing too much bias. It can also reduce factoring the outliers (noise) into the model and can reduce the model space (lasso).

2.

No. we did not use regularization. By using least squares we actually weighted the outliers (noise) heavier. We did not add any new information to our data to find a regression function.

3.

No. In a simple predictable dataset, a model can be created simply, and would not require regularization. Regularization is used to reduce the randomness, but if the data doesn’t have that, it is not necessary, and otherwise complicates the solution.

4.

L1 (Lasso) and L2 (ridge) regularization methods.