**Part 1**

1. Matlab function was used to build linear model: mvregress([ones(1000,1),X1],Y1);  
   a column of ones were added for computing intercept.   
   This linear regression gives an R-squared value of 0.3742, meaning ~37% of variance can be explained by the linear model.  
   The residuals are plotted in the following figure. The random distribution of the residual terms indicates a good fit, and that this linear model is appropriate for this data.  
   
2. Matlab lasso function was used for lasso regression with lambda varying from 0 to 0.3 at a step size of 0.001. With increasing lambda, lasso regression was able to reduce coefficiencies to zero, meaning dropping some dimensions (following figure). During this process, we will be able to find which dimensions contribute more to Y than others, and we can select particular dimensions based on the desired number of nonzero coefficiencies.   
     
   Different numbers of feature were selected based on lasso regression result and used for fitting a linear model. 10-fold cross validation was used to test how good the fitting was. The following figure illustrates the training and validation errors.   
     
     
     
   Based on the cross validation result, adding first 4 features decrease error drastically, which indicates that the first 4 features are very important. However, adding more features do decrease validation error slowly, indicating a secondary role of these features. Depsite this, the last 4 features do not seem play an important role in explaining the labels, so they may be just purely noise.

Part II

1. Linear model for X2 and Y2.   
     
   Errors are plotted in the following figure. Apparently the residuals are not randomly distributed, suggesting that linear model is not appropriate for this problem.   
   
2. Piece-wise linear model. Two methods are used here: discrete piece-wise linear model and continuous piece-wise linear model. In both models, two knots were chosen at 0.3 and 0.8.  
   **A**. discrete piece-wise: data were divided to three parts (0<=x<0.3, 0.3<=x<0.8, 0.8<=x<=1) and each part was used for linear regression.   
   Errors are plotted in the following figure. In this case, errors are randomly distributed, indicating a decent fitting.   
     
   B. Continuous piece-wise linear model: data were still divided to three parts (0<=x<0.3, 0.3<=x<0.8, 0.8<=x<=1). The modeling function is:  
   Briefly, two additional features were added to the original linear model. Solving this linear equation gives us a continuous piece-wise linear model illustrated in the following figure.   
     
     
   The errors are plotted in the following figure. Residuals also have a random distribution pattern, indicating a decent fitting.   
   
3. Additive piece-wise linear model means that each basis function only depends on individual feature/dimension. X3 has two dimensions (denoted as X1 and X2), therefore the overall modeling function is:  
   Each dimension has a set of basis functions, and the overall function is a linear combination of the two sets.   
   In this particular problem, both X1 and X2 vary from 0 to 5, giving us 5 sections for each dimension. Therefore the detailed modeling function is:   
   The intercept was combined into one , and there are totally 10 coefficiencies besides the intercept.   
   Real data is plotted as a heatmap:  
     
   The fitted linear function is illustrated in the following picture (each axis represents one dimension, colors represent theoretical value for that point):  
     
   The errors are plotted in the following figure.   
   

(Rough codes can be found on the website: http://www.unc.edu/~zyu/ml134/midterm/midterm.html)