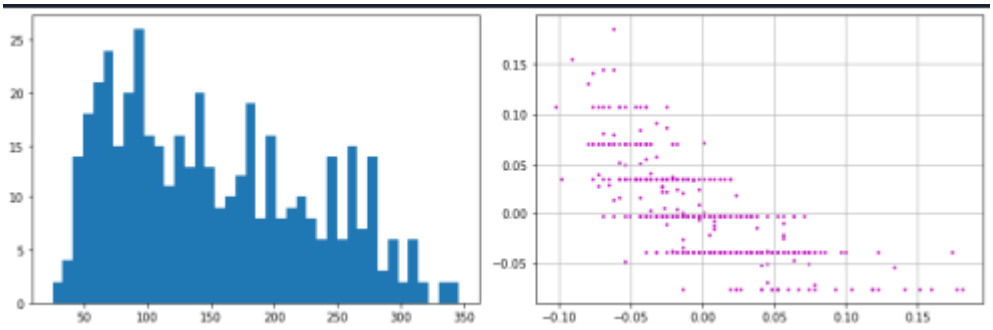
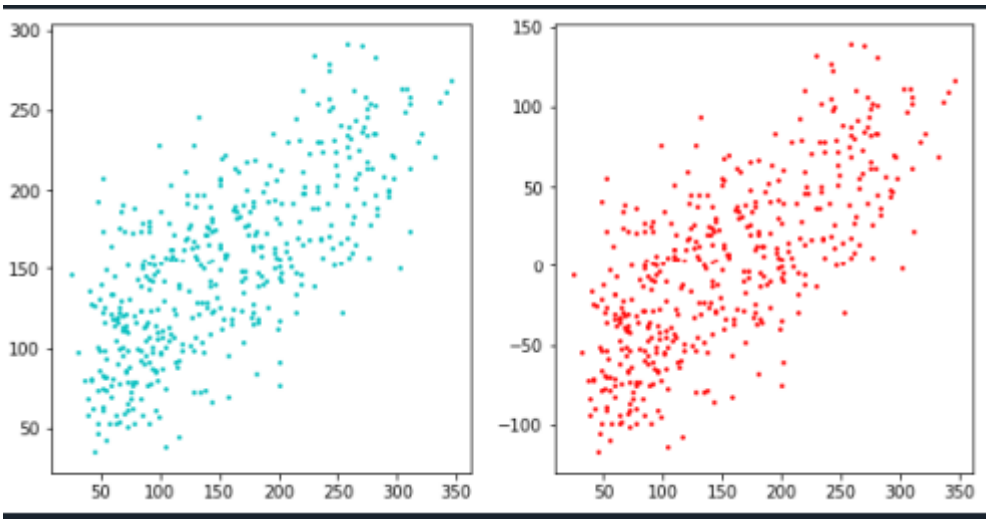
Name: Mao YANG
Email:my4n20@soton.ac.uk

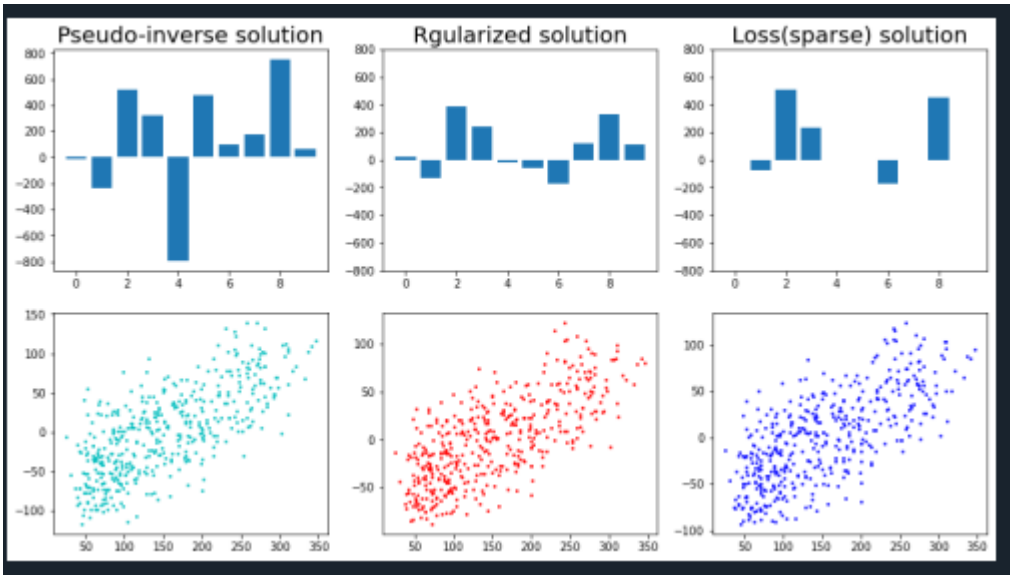
1.Linear regression on diabetes dataset:



2.Comparing pseudo-inverse solution to sklearn output:

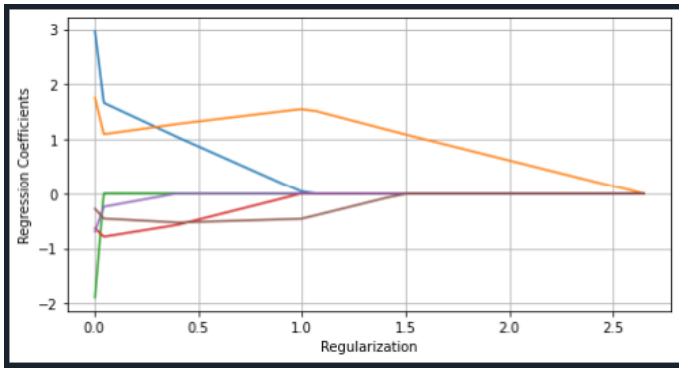


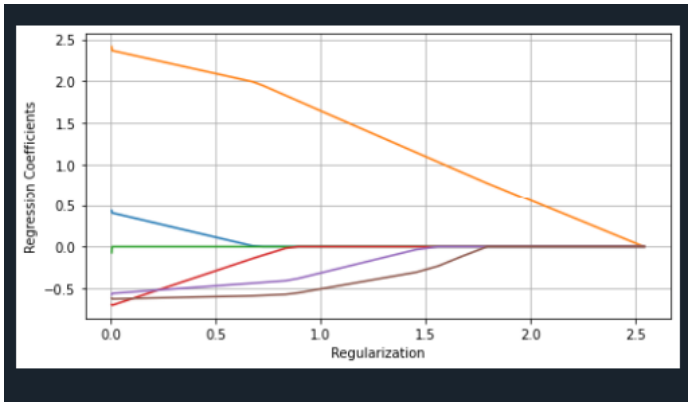
3.Tikhanov (quadratic) Regularizer:and and 4. Sparsity inducing (lasso) regularizer



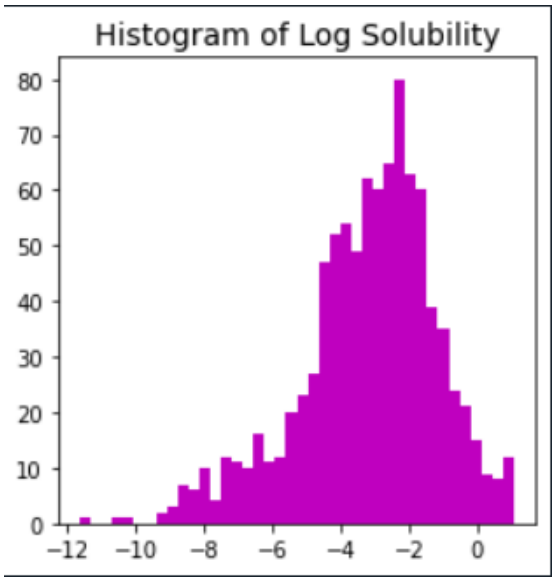
5. Lasso Regularization path on a synthetic example (Set up data):

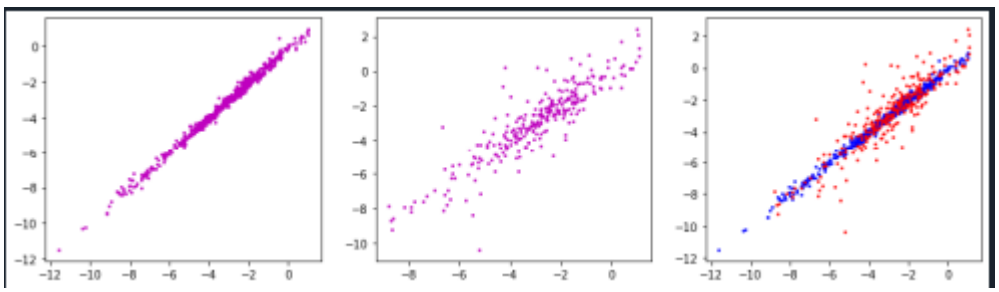
And 6. Lasso Regularization path on a synthetic example (Regression and paths):





7. Predicting Solubility of Chemical Compounds:





1. Load the data, split into training and test sets, implement a linear regression and plot the predicted solubilities against the true solubilities on the training and test sets. To facilitate

comparison, draw the two scatter plots side by side to the same scale on both axes.

2. Implement a lasso regularized solution and plot graphs of how the prediction error (on the test data) and the corresponding number of non-zero coefficients change with increasing regularization.

3. If you were to select the top ten features to predict solubility, what would they be? How good is the prediction accuracy with these slected features when compared to using all the features and a quadratic regularizer?

Are you able to make any comment comparing your results to those claimed in [3] or [4]