**Assignment 5 plus**

1. **Gradient Descent, Linear Regression, and Logistic Regression** (20 pts)

The **data\_hw5\_1.mat** file has three columns. The first column is x (input), the second column is y (output, a continuous one) and the third one z (output, a discrete one). We want to build a linear regression model for y as a function of x, and a logistic regression model for z as a function of x.

* 1. Use the gradient descent technique to find parameters of the linear regression model (you need to minimize RSS)
  2. Use the gradient descent technique to find parameters of the logistic regression model (you need to minimize NLL).
  3. Use Matlab (fitglm) or Python toolbox to find the linear regression model parameters and compare it with your result in step **a**
  4. Use Matlab (fitglm) or Python toolbox to find the logistic regression model parameters and compare it with your result in step **b**

**2** **Model Selection, Linear Regression** (56 pts)

The **data\_hw5\_2.mat** file has 6 columns. The first 6 columns are the input (X) and the last column is output (y). We are interested in building the best model which gives the lowest generalized error.

1. Build a linear regression model and show the model parameters. Show the performance metrics including RSS, BIC, and R^2.
2. Use the best subset selection method to find the best model. Provide performance metrics and model parameters, similar to step **a**. Discuss which approach you took for the model comparison – BIC, cross-validation, etc.
3. Use the forward stepwise model selection method to find the best model. Provide performance metrics and model parameters, similar to step **a**.
4. Use the backward stepwise model selection method to find the best model. Provide performance metrics and model parameters, similar to step **a**.
5. Use the ridge regression method and discuss the model selection result. Provide model parameters similar to step a.
6. Use the lasso regression method and discuss the model selection result. Provide model parameters similar to step a.
7. Compare the performance result of the models you picked in steps **a** to **f** and suggest which one you pick.

**3** **Model Selection, Logistic Regression** (24 pts)

The **data\_hw5\_3.mat** file has 3 columns. The first 2 columns are the input (X) and the last column is output (y, binary). We are interested in building the best model which gives the best precision or prediction accuracy. Here, we are interested in checking higher-order terms of the input including , , and and whether adding these terms will improve model accuracy. The model input will be , , , , and – you can also add higher-order terms.

1. Build a logistic regression model and show the model parameters. Show the performance metrics including BIC and precision.
2. Add higher-order terms, and build a logistic regression model and show the model parameters. Show the performance metrics including BIC and precision.
3. Pick a model selection strategy and find the model with a minimum number of predictor which gives the best performance. Discuss your methodology and model performance
4. Discuss three models you built in steps **a** to **c**.