# Homework #2

**Assigned: 02/09/2021**
**Due: 02/23/2021 (11:59 PM on CANVAS)**

## A. Theory Problems

1. **Linear Regression for one Variable:** Consider the problem of a single input variable $x$ and output variable $y$. Assume we want to fit a simple linear model (without an intercept) of the form $\hat{y} = \theta x$, and accordingly find the best value of $\theta$. Assume we have $N$ measurements of the input and output where the $n$-th measurement is given by $x_n$ and $y_n$ with corresponding estimate $\hat{y}_n = \theta x_n$.

   To find the best model, we want to minimize the means square error (MSE), $MSE = \frac{1}{N}\sum_{n=1}^{N}(\hat{y}_n - y_n)^2$.

   a) Find the derivative of the mean square error with respect to $\theta$, $\frac{\partial MSE}{\partial \theta}$.
   b) Using your answer from (a), the find the optimum value of $\theta$. **Hint: What happens to the derivative of a function at a minimum?*

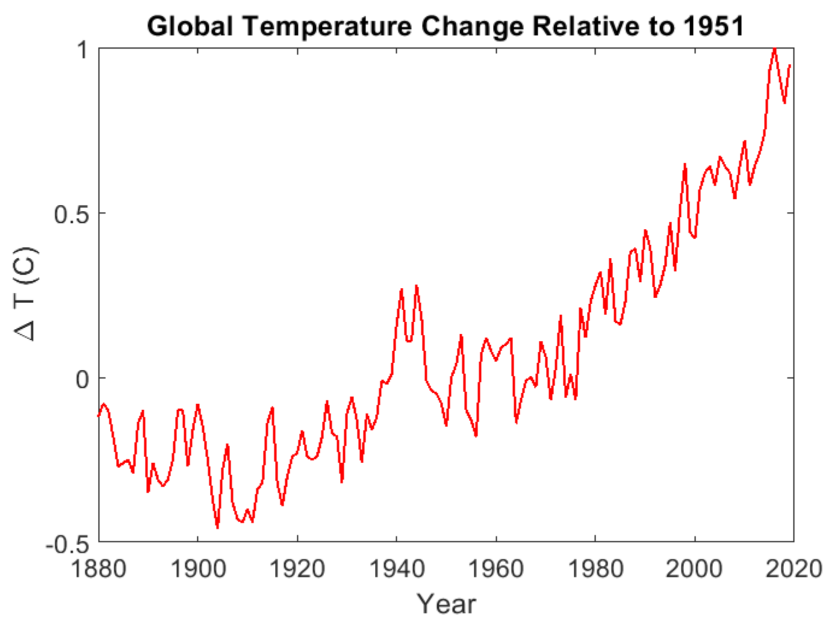
2. **Derivation of the Normal Equations:** In problem 1, you determined the best linear model for a single variable; for this problem you will find the best linear model for multi-linear regression. Consider the problem of $K$ input variables such that the input vector is given by $\boldsymbol{x} = [x_1, x_2, \dots x_{K-1}, x_K]$. Assume we want to fit a multi-linear model (without an intercept) of the form $\hat{y} = \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_3 + \cdots \theta_K x_K = \boldsymbol{\theta}^{\boldsymbol{T}}\boldsymbol{x}$, where $\boldsymbol{\theta} = [\theta_1, \theta_2, \dots \theta_{K-1}, \theta_K]^T$. Assume we have $N$ measurements of the input and output where the $n$-th measurement is given by $\boldsymbol{x}^{(n)}$ and $y^{(n)}$ with corresponding estimate $\hat{y}^{(n)} = \boldsymbol{\theta}^{\boldsymbol{T}}\boldsymbol{x}^{(n)}$. We will also define the output measurement vector as $\boldsymbol{y} = [y_1, y_2, \dots y_{N-1}, y_N]$.

   To find the best model, we want to minimize the means square error (MSE), $MSE = \frac{1}{N}\sum_{n=1}^{N}\left(y^{(n)} - \hat{y}^{(n)}\right)^2$.

   a) Show that the vector $\hat{\boldsymbol{y}} = [\hat{y}_1, \hat{y}_2, \dots \hat{y}_{N-1}, \hat{y}_N]^T$, can be written as $\hat{\boldsymbol{y}} = \boldsymbol{X\theta}$**,** where $\boldsymbol{X}$ is a matrix. Explicitly show the rows of the matrix $\boldsymbol{X}$**.**
   b) Show that the MSE can be written as $\boldsymbol{MSE} = |\boldsymbol{y} - \boldsymbol{X\theta}|^{\boldsymbol{2}}$ where the "| |" operator corresponds to the magnitude of a vector.
   c) Find an expression for the optimum model parameter vector $\boldsymbol{\theta}$**.** *Hint*: Use the fact that the gradient of the MSE is given by* $\frac{\partial}{\partial \boldsymbol{\theta}}|\boldsymbol{y} - \boldsymbol{X\theta}|^2 = -2\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{y} + 2\boldsymbol{X}^{\boldsymbol{T}}\boldsymbol{X\theta}$**.**

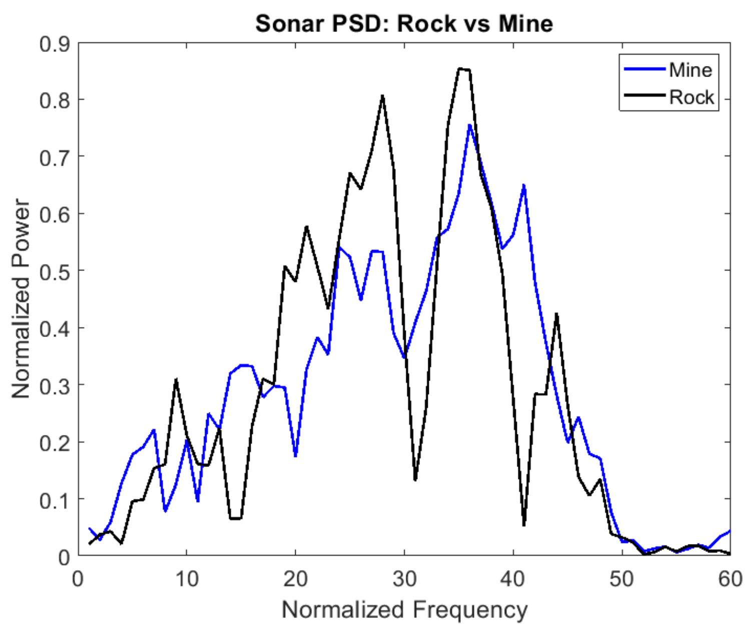
# B. Coding Problems

3. **Polynomial Regression of Temperature Data:** For this problem you will perform polynomial regression of global temperature data from 1880-2020. The datafile 'global_temp_data.csv' contains the year in the first column and yearly average temperature (change from temperature in 1951) in the second column. The figure below shows a plot of the global temperature change as a function of time.




a) Using a 10-fold cross validation (CV) method, build the optimal polynomial model to fit the temperature data. What is the degree of the polynomial? Include a plot of the mean CV score as a function of the polynomial degree (up to 20).
b) Overly a plot of your model over the data between 1880 and 2040.
c) Using the model, what is the expected temperature change in 2040?
d) Provide an estimate of the error (in degrees celcius) in your prediction from (c). *Hint* What error metric could be useful?*

4. **Mine or Rock Classification using Sonar Data:** In this problem you will use data form sonar measurements to distinguish between two classes of objects: "Rock" or "Mine". The datafile 'sonar.csv' contains sonar signatures for 208 labeled scenarios. Each row corresponds to a single signature contain 60 power values corresponding to 60 frequency bins. The $61^{st}$ column of each row correspond to the class of that signature, 'M' for mine and 'R' for rocks. An example of each signature is shown in the figure below.




Split the data using an 80-20 split of testing to training data. Using the training data, build a classifier model using:

a) Perceptron
b) Logistic Regression
c) Linear Support Vector Machine
d) Polynomial Support Vector Machine

Run of each of the above models on the testing data and show a confusion matrix and an ROC curve for each. *Which model is the best option for this dataset?*