Assignment 2, Linear Regression

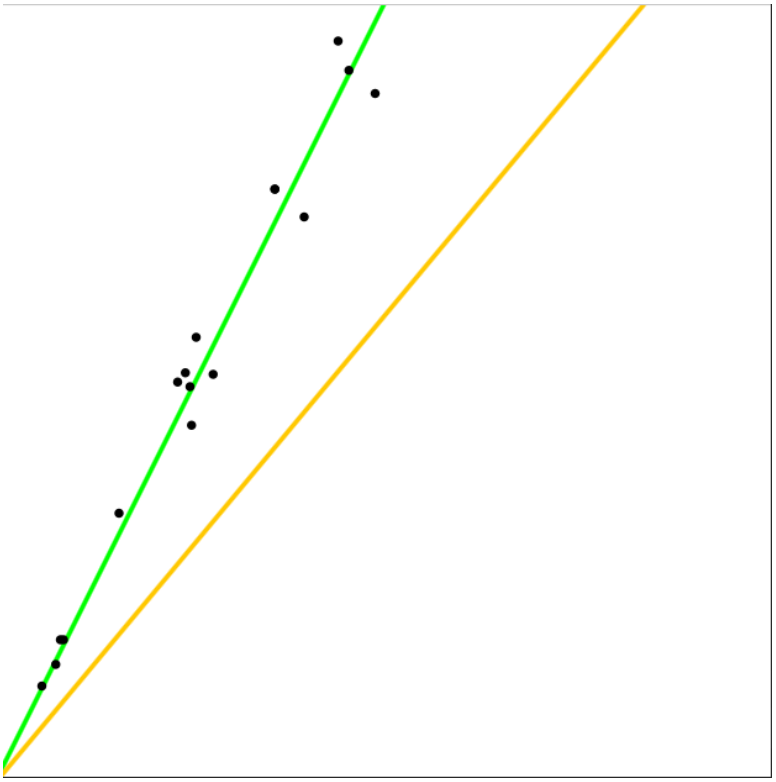
CSC736 Machine Learning
Spring 2022
Max Score: 100

Objectives

- Learn and demonstrate knowledge of Linear Regression.
- Visualize the learning process of regression.

A linear regressor is able to predict values with given inputs based on provided training dataset. In this assignment, you are required to develop a program that is able to:

1. Generate points in the training set.
   - Arbitrarily define a line $y = wx + b$ (eg. y = 2x+3) as your ground truth line;
   - Generate 20 random data points (randomly select 20 x and calculate 20 y accordingly) from the line defined above. The y value on each point needs to randomly add or minus a noise with the range of $10\% * y$. For example, assuming your line is $y = 2x + 3$. Your first point is $x = 10$, $y = 23 + rand(23 * 0.1)$ or $y = 23 - rand(23 * 0.1)$.
   - Visualize the line in green and the 20 points (filled circles) on a graphic user interface similar to the attached figure.
2. Implement a linear regression with the gradient descent learning algorithm.
   - Randomly initialize the weight and bias to a double within (0, 1).
   - Set your learning rate $\eta = 0.000001$
   - Train your linear regressor by the gradient descent learning algorithm with the provided training data generated from the previous step.
     - Define "epoch" as one iteration of training all 20 points one time.
     - For each epoch, iterating through 20 training points $(\hat{x}_i, \hat{y}_i)$:
       * Use your linear model to calculate $y_i = w\hat{x}_i + b$
       * Calculate the accumulated error for bias: $err_b = \frac{1}{m}\sum_{i=1}^{m}(\hat{y}_i - y_i)$
       * Calculate the accumulated error for weight: $err_w = \frac{1}{m}\sum_{i=1}^{m}[(\hat{y} - y_i) * \hat{x}_i]$



  - Using this formula to update the bias: $b = b + \eta \times error_b$
  - Using this formula to update the weight: $w = w + \eta \times error_w$
- Visualize the line represented by the current weights at the end of each epoch on GUI. (like an animation.)
- Output the number of mean square error ($mse = \frac{1}{m} \sum_{i=1}^{m} (\hat{y}_i - y_i)^2$) on the training data at the end of each epoch.
- Train your linear regression model for at least 500 epochs.