CS613 Machine Learning - HW 2 - Linear Regression

McWelling Todman

Due: February 15^{th}, 2019

1 Theory

1. (10pts) Consider the following data:

$$A = \begin{bmatrix} -2 & 1 \\ -5 & -4 \\ -3 & 1 \\ 0 & 3 \\ -8 & 11 \\ -2 & 5 \\ 1 & 0 \\ 5 & -1 \\ -1 & -3 \\ 6 & 1 \end{bmatrix}$$

(a) Compute the coefficients for the linear regression using least squares estimate (LSE), where the second value (column) is the dependent variable (the value to be predicted) and the first column is the sole feature. Show your work and remember to add a bias feature and to standardize the features. Compute this model using **all** of the data (don't worry about separating into training and testing sets).

Our first step is to standardize the component feature X of our data in matrix A. In order to do this, we will need to calculate the mean (μ_i) and standard deviation (σ_i) for each feature:

$$\mu_X\colon \frac{\Sigma(X_i)}{10} = -.9$$

$$\mu_Y\colon \frac{\Sigma(Y_i)}{10} = 1.4$$

$$\sigma_X\colon \sqrt{\frac{\Sigma(X_i - \mu_x)}{10}} = 4.2282 \qquad \sigma_Y\colon \sqrt{\frac{\Sigma(y_i - \mu_y)}{10}} = 4.2740$$

We are now able to standardize our explanatory variable A_X by applying the following logic:

$$\text{Standardization } A_x\text{: for all } x_i \in A_x \longrightarrow \frac{x_i - \mu_x}{\sigma_x}$$

$$\begin{bmatrix} -0.2602 & 1 \\ -0.9697 & -4 \\ -0.4967 & 1 \\ 0.2129 & 3 \\ -1.6792 & 11 \\ -0.2602 & 5 \\ 0.4494 & 0 \\ 1.3954 & -1 \\ -0.0237 & -3 \\ 1.6319 & 1 \end{bmatrix}$$

After standardizing, we add an additional feature with a value of one to the data. We call this additional feature the bias (Θ_0), and it is considered an expansion of our set of explanatory variables X:

$$\begin{bmatrix} 1 & -0.2602 & 1 \\ 1 & -0.9697 & -4 \\ 1 & -0.4967 & 1 \\ 1 & 0.2129 & 3 \\ 1 & -1.6792 & 11 \\ 1 & -0.2602 & 5 \\ 1 & 0.4494 & 0 \\ 1 & 1.3954 & -1 \\ 1 & -0.0237 & -3 \\ 1 & 1.6319 & 1 \end{bmatrix}$$

Once our bias feature has been added, we are free to compute the vector containing the weights of our model using the formula $\Theta = (X^T X)^{-1} X^T Y$:

1. $\Theta = (X^T X)^{-1} X^T Y$

2a. $(X^T X)^{-1} = ($
$$\begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -0.2602 & -0.9697 & -0.4967 & 0.2129 & -1.6792 & -0.2602 & 0.4494 & 1.3954 & -0.0237 & 1.6319 \end{bmatrix}$$
$$\begin{bmatrix} 1 & -0.2602 \\ 1 & -0.9697 \\ 1 & -0.4967 \\ 1 & 0.2129 \\ 1 & -1.6792 \\ 1 & -0.2602 \\ 1 & 0.4494 \\ 1 & 1.3954 \\ 1 & -0.0237 \\ 1 & 1.6319 \end{bmatrix})^{-1} = \begin{bmatrix} 0.1000 & 0.0000 \\ 0.0000 & 0.1111 \end{bmatrix}$$

2b. $X^T Y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -0.2602 & -0.9697 & -0.4967 & 0.2129 & -1.6792 & -0.2602 & 0.4494 & 1.3954 & -0.0237 & 1.6319 \end{bmatrix}$
$$\begin{bmatrix} 1 \\ -4 \\ 1 \\ 3 \\ 11 \\ 5 \\ 0 \\ -1 \\ -3 \\ 1 \end{bmatrix} = \begin{bmatrix} 14.0000 \\ -15.7040 \end{bmatrix}$$

3. $\Theta = (X^T X)^{-1} X^T Y = \begin{bmatrix} 0.1000 & 0.0000 \\ 0.0000 & 0.1111 \end{bmatrix} \begin{bmatrix} 14.0000 \\ -15.7040 \end{bmatrix} = \begin{bmatrix} 1.4000 \\ -1.7449 \end{bmatrix}$

Having computed our coefficient vector (Θ), we are able to express our model as:

$$\hat{Y} = 1.4000 - 1.7449X$$

2. For the function $g(x) = (x-1)^4$, where x is a single value (not a vector or matrix):

   (a) (3pts) What is the gradient with respect to x? Show your work to support your answer.

   1. $g(x) = (x-1)^4$

   2. $\frac{dg}{dx} = 4(x-1)^3$ by the chain rule

   (b) (3pts) What is the global minima for $g(x)$? Show your work to support your answer.

   1. $\frac{dg}{dx} = 4(x-1)^3 = 0$

   2. $4(1-1)^3 = 0$

   3. global minima occurs when $x = 1, (g(x) = 0)$

   (c) (3pts) Plot x vs $g(x)$ use a software package of your choosing.

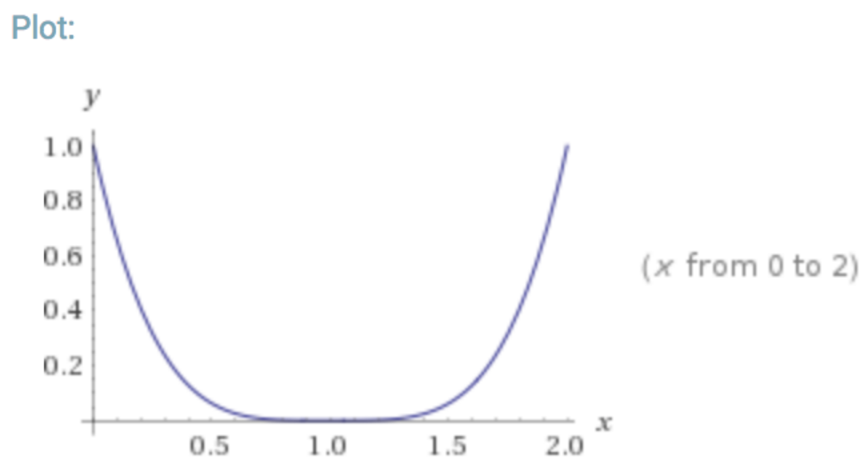

Figure 1: Plot of $g(x) = (x-1)^4$

2 Closed-form Linear Regression

1. Final Model

$$y = \theta_0 + \theta_1 x_1 + ...$$

$$\hat{Y} = 3044.67 + 1008.17x_1 - 298.86x_2$$

2. RMSE

$$\text{RMSE} = 572.6601$$

3 S-folds Validation

1. The average and standard deviation of the root mean squared error for $S = 3$ over the 20 different seed values..

$$\text{Average RMSE} = 618.9162$$

$$\text{Standard Deviation RMSE} = 30.2128$$

2. The average and standard deviation of the root mean squared error for $S = 5$ over the 20 different seed values.

$$\text{Average RMSE} = 614.1111$$

$$\text{Standard Deviation RMSE} = 24.9743$$

3. The average and standard deviation of the root mean squared error for $S = 20$ over 20 different seed values.

$$\text{Average RMSE} = 607.1958$$

$$\text{Standard Deviation RMSE} = 6.7973$$

4. The average and standard deviation of the root mean squared error for $S = N$ (where N is the number of samples) over 20 different seed values. This is basically *leave-one-out* cross-validation.

$$\text{Average RMSE} = 540.5283$$

$$\text{Standard Deviation RMSE} = 372.4688$$

4 Local Regression

1. RMSE

$$\text{RMSE} = 353.7712$$

5 Gradient Descent

1. Final Model

$$y = \theta_0 + \theta_1 x_1 + ...$$

$$\hat{Y} = 3044.13 + 1007.48x_1 - 298.26x_2$$

2. RMSE

$$\text{RMSE} = 554.87$$

3. Plot of RMSE for Training and Testing Data vs Gradient Descent iteration number

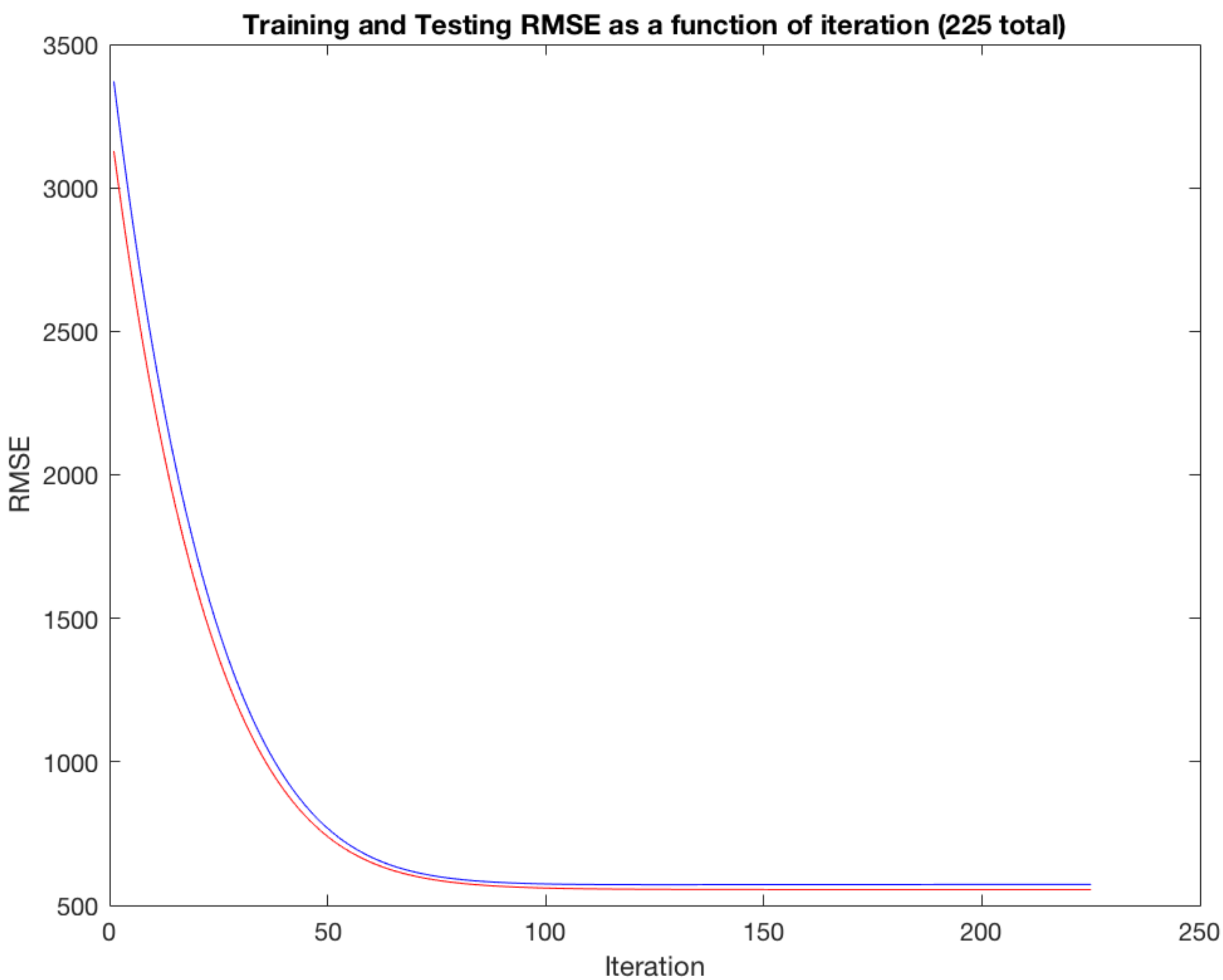

Figure 2: RMSE by iteration (train = red, test = blue)