ER90C, Data, Environment and Society

Fall 2018

Lab # 7

Distributed: October 8, 2018

Due: in class.

1. Suppose you want to fit the model $y_i = \beta_0 + \beta_1 x_i + \epsilon_i$ to the data below. Draw your best guess for what the model would be: (a) using a solid line to represent the model you think a mean squared error loss function would give and (b) a dashed line for your guess at what you'd get with a a mean absolute error loss function. Explain why you drew the lines as you did.

The MSE line has a larger slope because it is more influenced by the extreme value at $(x, y) = (1.5, 6)$.

2. Consider the following data set. It shows how much people spent to send a package via the postal service. The data also include the time zone they sent the package to and its weight. Write down a model that could be used to predict shipping cost from the remaining data. You may need to draw addtional columns in the dataframe to 'code' categorical variables as we did in class.

Name	Cost	Zone
Lars	$28	Western
Maneesh	$22	Mountain
Goodwill	$4	Central
Ruzena	$15	Eastern
Claudio	$21	Western
Magali	$8	Mountain
Tsu-Jae	$18	Central
Malebogo	$16	Eastern
Rongxin	$15	Other
Maria	$13	Other

Name	Cost	Zone	Package weight (lb)	x_western	x_mountain	x_central	x_eastern
Lars	$28	Western	33	1	0	0	0
Maneesh	$22	Mountain	34	0	1	0	0
Goodwill	$4	Central	1	0	0	1	0
Ruzena	$15	Eastern	25	0	0	0	1
Claudio	$21	Western	23	1	0	0	0
Magali	$8	Mountain	12	0	1	0	0
Tsu-Jae	$18	Central	6	0	0	1	0
Malebogo	$16	Eastern	17	0	0	0	1
Rongxin	$15	Other	60	0	0	0	0
Maria	$13	Other	2	0	0	0	0

$$\text{cost} = \beta_0 + \beta_1 \times \text{weight} + \beta_2 x_{western} + \beta_3 x_{mountain} + \beta_4 x_{central} + \beta_5 x_{eastern}$$