Stat637 Homework 7 [1]

Arthur Lui

11 March 2015

1 Proportional Odds Model

$$log\left(\frac{p_1 + ... + p_j}{p_{j+1} + ... + p5}\right) = \beta_{0j} + \beta_1\ spline(helpProportion) + \beta_2\ year + \beta_3\ spline(negativeWords),$$

where $j = 1, 2, 3, 4$, *helpProportion = help.yes/help.count*, and *negativeWords* is defined to be the count of words with negative sentiments in a review. These words were identified and selected by after reading the reviews in the training set. The following words were selected:

too, nothing, why, how, shortcoming, didn't like, don't like, difficult, stinks, sucks, blows, annoy, bad, not recommend

2 Parameters

	Estimate	Std. Error	z value
(Intercept):1	-157.07	143.32	-1.10
(Intercept):2	-155.53	143.30	-1.09
(Intercept):3	-154.24	143.27	-1.08
(Intercept):4	-153.34	143.26	-1.07
bs(helpful.prop)1	7.60	2.75	2.76
bs(helpful.prop)2	-6.12	2.63	-2.33
bs(helpful.prop)3	2.24	1.15	1.95
year	0.08	0.07	1.07
bs(neg.count)1	7.43	2.86	2.60
bs(neg.count)2	-3.49	9.21	-0.38
bs(neg.count)3	1.89	1.86	1.02

Table 1: Model Summary

3 Prediction Error

From Table 2, we can compute the prediction accuracy to be (1+3+8)/(1+1+2+3+1+1+2+8) = 12/19 = 63%. The MSE was computed to be 0.526.

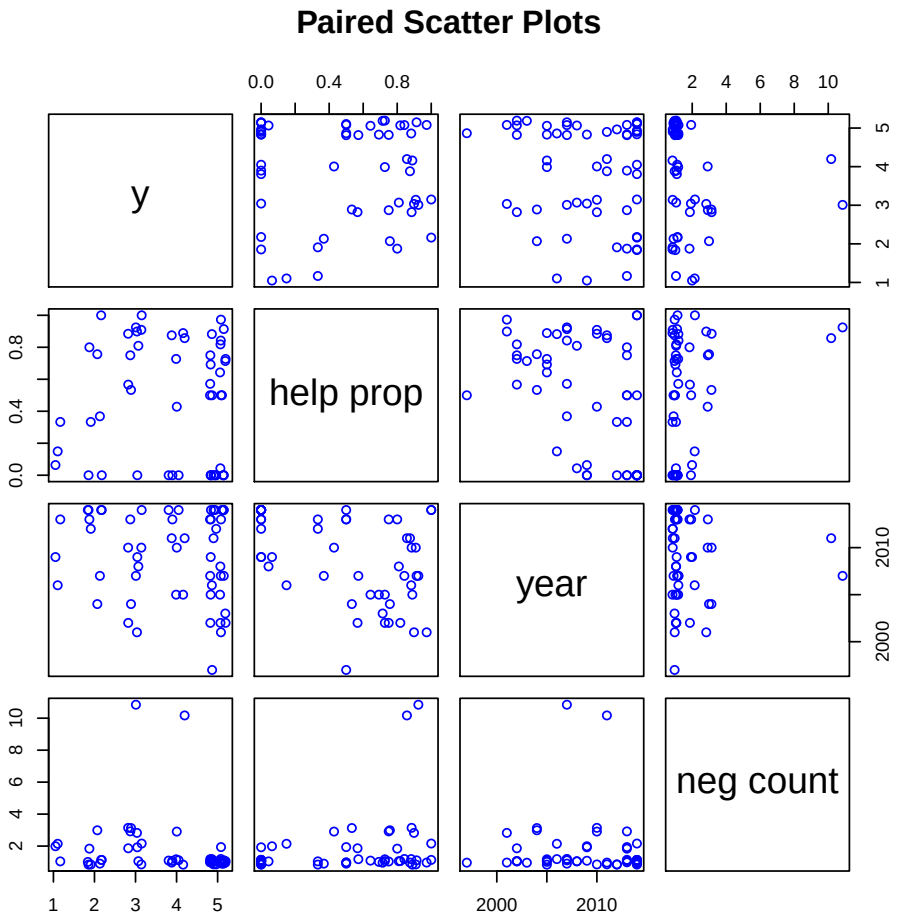
4 Plots

The paired scatter plot below shows the relationship between the covariates chosen and the response variable. A strong trend cannot be seen between the covariates and the response.

[1] https://github.com/luiarthur/Fall2014/blob/master/Stat637/competition

	1	3	5
1	1	0	0
2	1	2	0
3	0	3	1
4	0	1	2
5	0	0	8

Table 2: Confusion Matrix. Rows are actual responses, columns are predicted responses



5 Value of Model

The model is not easy to interpret. Nevertheless, it predicts well, with an MSE of 0.526 and few parameters.