Classification: KNN vs. LS

Mingliang Wang

1 KNN and LS for classification

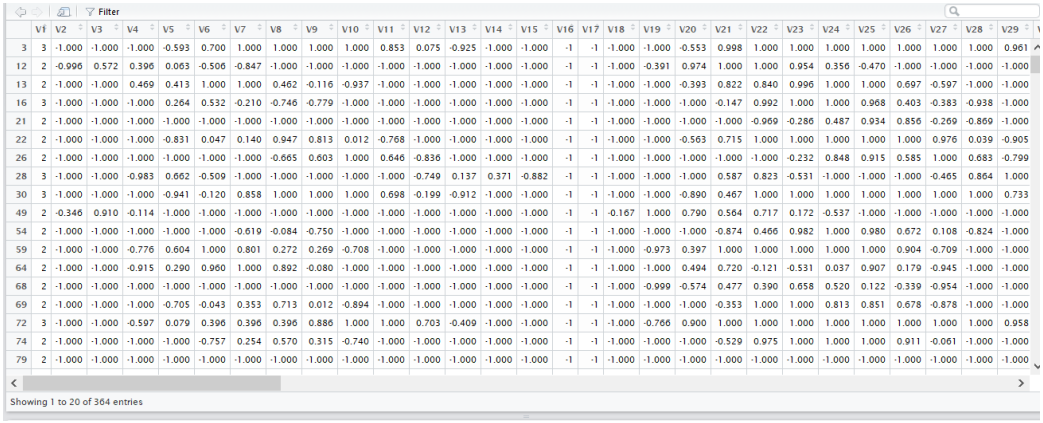
Compare the classification performance of linear regression and k nearest neighbor classification on the zipcode data. In particular, we consider only the category 2's and 3's. For KNN, we studied the cases that $k = 1$, 3, 5, 7 and 15. Show both the training and test error for each choice. The least square (LS) method make a decision like this: if prediction > 2.5, class $= 3$, else class $= 2$.

2 Data description

The data is stored in a file named "zipdata". For training data, every column stands for a covariate and "%" stands for the response in train%.txt; for the test data set, the first column stands for the response and other columns stands for covariates. A part of the training data is given in Figure 1.

3 Results

The KNN and LS are implemented using R and the script is given in "KNNvsLS.R". The results are plot in Figure 2 in which the training and testing errors for both methods are presented. The training error of KNN increase with k and so as the classification error. In addition, KNN performance better than the LS at all k values in testing. So we would prefer the KNN for the classification.



	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	V18	V19	V20	V21	V22	V23	V24	V25	V26	V27	V28	V29
3	3	-1.000	-1.000	-1.000	-0.593	0.700	1.000	1.000	1.000	1.000	0.853	0.075	-0.925	-1.000	-1.000	-1	-1	-1.000	-1.000	-0.553	0.998	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.961
12	2	-0.996	0.572	0.396	0.063	-0.506	-0.847	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1	-1	-1.000	-0.391	0.974	1.000	1.000	0.954	0.356	-0.470	-1.000	-1.000	-1.000	-1.000
13	2	-1.000	-1.000	0.469	0.413	1.000	1.000	0.462	-0.116	-0.937	-1.000	-1.000	-1.000	-1.000	-1.000	-1	-1	-1.000	-1.000	-0.393	0.822	0.840	0.996	1.000	1.000	0.697	-0.597	-1.000	-1.000
16	3	-1.000	-1.000	-1.000	0.264	0.532	-0.210	-0.746	-0.779	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1	-1	-1.000	-1.000	-1.000	-0.147	0.992	1.000	1.000	0.968	0.403	-0.383	-0.938	-1.000
21	2	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1	-1	-1.000	-1.000	-1.000	-1.000	-0.969	-0.286	0.487	0.934	0.856	-0.269	-0.869	-1.000
22	2	-1.000	-1.000	-1.000	-0.831	0.047	0.140	0.947	0.813	0.012	-0.768	-1.000	-1.000	-1.000	-1.000	-1	-1	-1.000	-1.000	-0.563	0.715	1.000	1.000	1.000	1.000	1.000	0.976	0.039	-0.905
26	2	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-0.665	0.603	1.000	0.646	-0.836	-1.000	-1.000	-1.000	-1	-1	-1.000	-1.000	-1.000	-1.000	-1.000	-0.232	0.848	0.915	0.585	1.000	0.683	-0.799
28	3	-1.000	-1.000	-0.983	0.662	-0.509	-1.000	-1.000	-1.000	-1.000	-1.000	-0.749	0.137	0.371	-0.882	-1	-1	-1.000	-1.000	-1.000	0.587	0.823	-0.531	-1.000	-1.000	-1.000	-0.465	0.864	1.000
30	3	-1.000	-1.000	-1.000	-0.941	-0.120	0.858	1.000	1.000	1.000	0.698	-0.199	-0.912	-1.000	-1.000	-1	-1	-1.000	-1.000	-0.890	0.467	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.733
49	2	-0.346	0.910	-0.114	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1	-1	-0.167	1.000	0.790	0.564	0.717	0.172	-0.537	-1.000	-1.000	-1.000	-1.000	-1.000
54	2	-1.000	-1.000	-1.000	-1.000	-1.000	-0.619	-0.084	-0.750	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1	-1	-1.000	-1.000	-1.000	-0.874	0.466	0.982	1.000	0.980	0.672	0.108	-0.824	-1.000
59	2	-1.000	-1.000	-0.776	0.604	1.000	0.801	0.272	0.269	-0.708	-1.000	-1.000	-1.000	-1.000	-1.000	-1	-1	-1.000	-0.973	0.397	1.000	1.000	1.000	1.000	1.000	0.904	-0.709	-1.000	-1.000
64	2	-1.000	-1.000	-0.915	0.290	0.960	1.000	0.892	-0.080	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1	-1	-1.000	-1.000	0.494	0.720	-0.121	-0.531	0.037	0.907	0.179	-0.945	-1.000	-1.000
68	2	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1	-1	-1.000	-0.999	-0.574	0.477	0.390	0.658	0.520	0.122	-0.339	-0.954	-1.000	-1.000
69	2	-1.000	-1.000	-1.000	-0.705	-0.043	0.353	0.713	0.012	-0.894	-1.000	-1.000	-1.000	-1.000	-1.000	-1	-1	-1.000	-1.000	-1.000	-0.353	1.000	1.000	0.813	0.851	0.678	-0.878	-1.000	-1.000
72	3	-1.000	-1.000	-0.597	0.079	0.396	0.396	0.396	0.886	1.000	1.000	0.703	-0.409	-1.000	-1.000	-1	-1	-1.000	-0.766	0.900	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.958
74	2	-1.000	-1.000	-1.000	-1.000	-0.757	0.254	0.570	0.315	-0.740	-1.000	-1.000	-1.000	-1.000	-1.000	-1	-1	-1.000	-1.000	-1.000	-0.529	0.975	1.000	1.000	1.000	0.911	-0.061	-1.000	-1.000
79	2	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1	-1	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000	-1.000

Showing 1 to 20 of 364 entries

Figure 1: zipdata description

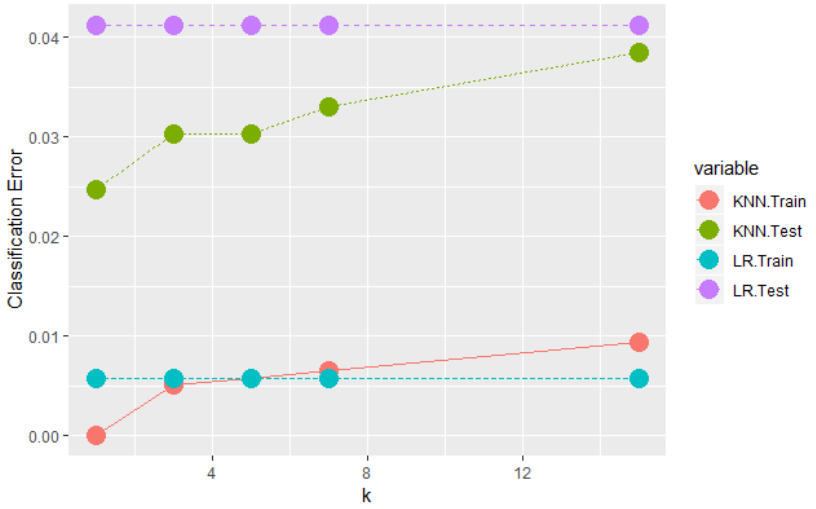


Figure 2: Classification error for KNN and LS