

**CS 498 Additional Homework for 4 credits (kakilai2, tlu16)**

**Result is shown below for all the 35 data sets:**

<b>Data Set</b>	<b>Best Regularization Constant (min lambda)</b>	<b>Average cv accuracy</b>	<b>Misclassification rate</b>	<b>Number of genes the best model used</b>
1	0.065017982	0.958333333	0.013888889	11
2	0.01311748	0.930555556	0.041666667	42
3	0.041123777	0.935960591	0.019704433	44
4	0.11021124	0.971153846	0.019230769	7
5	0.007532924	0.625	0.025	35
6	0.019254086	0.958333333	0.013888889	19
7	0.010603193	0.958333333	0.041666667	29
8	0.02336176	0.961325967	0.033149171	16
9	0.27271803	0.783783784	0.216216216	0
10	0.123783301	0.68	0.12	30
11	0.028904622	0.857142857	0	17
12	0.11700329	0.772727273	0	11
13	0.042002661	0.794117647	0.264705882	13
14	0.2062975	0.785714286	0.119047619	14
15	0.038888927	0.668421053	0.210526316	90
16	0.097737851	0.831168831	0.194805195	15
17	0.102104714	0.911764706	0.078431373	7
18	0.004454524	0.908045977	0	114
19	0.28245052	0.816326531	0.06122449	4
20	0.007416018	0.987903226	0	15
21	0.02063551	0.866935484	0.129032258	95
22	0.067288408	0.857142857	0	16
23	0.126103641	0.967741935	0	13
24	0.083844101	0.783333333	0.033333333	26
25	0.005240546	0.789473684	0	20
26	0.162372936	0.82	0.12	8
27	0.016283987	0.944134078	0	22
28	0.00591205	0.833333333	0.181818182	44
29	0.159387292	0.987951807	0	13
30	0.083789911	0.735294118	0.279411765	25
31	0.004168261	0.844036697	0	85
32	0.096315858	0.945945946	0	7
33	0.095479901	0.785714286	0	24
34	0.053065086	0.855769231	0.019230769	63
35	0.004750583	0.826086957	0	73

**Comment:**

- Regarding the good performance (low misclassification rate) of the original dataset, no transformation has been done for the data set (kernel boxcox) so there are no other parameters in the model apart from setting  $\alpha = 1$  for LASSO penalty.
- In certain data set, there are abnormal value ,i.e. character in numeric field, They are then converted to 'NA' value in the data set and are removed during the data cleaning process
- The misclassification rate is calculated by the proportion of misclassified observations by the model using *type.measure = "class"* in the prediction.