Supplementary File of Improved Space Forest: a Meta Ensemble Method

Mehmet Fatih Amasyali

October 15, 2017

1 Datasets

In Table A, the used datasets in the experiments are listed.

2 Detailed Accuracies of Improved Space and Original Versions of Ensembles

In tables B, C, and D, the average accuracies for each data set are given.
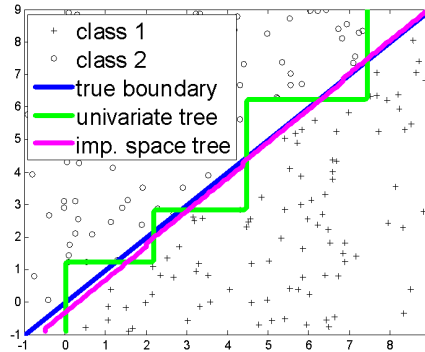
3 Figures

Figure A: The decision boundaries of the univariate and improved space trees over a linearly separated data set

Table A: Data set summary

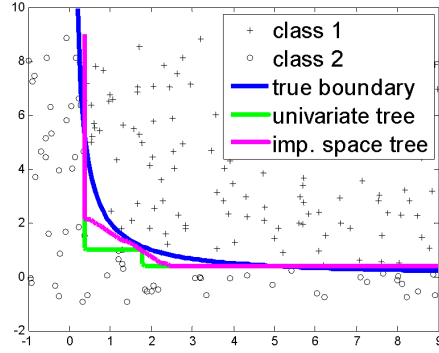| Data set ID	Data set	Features	Classes	Sample size	Data set ID	Data set	Features	Classes	Sample size
1	cmc	9	3	1473	22	diabetes	8	2	768
2	cnae	855	9	1080	23	glass	9	5	205
3	column3C	6	3	310	24	heart.sta.	13	2	270
4	ecoli	6	5	327	25	hepatitis	19	2	155
5	hillValley	100	2	606	26	ionosphere	33	2	351
6	monk	6	2	122	27	iris	4	3	150
7	move.Libras	90	15	360	28	labor	26	2	57
8	spambase	57	2	4601	29	letter	16	26	20000
9	transfusion	4	2	748	30	lymph	37	2	142
10	wineCult.	13	3	153	31	Prim.tumor	23	11	302
11	yeast	8	9	1479	32	ringnorm	20	2	7400
12	abalone	10	19	4153	33	segment	18	7	2310
13	audiology	69	5	169	34	sonar	60	2	208
14	autos	71	5	202	35	soybean	83	18	675
15	bal.scale	4	3	625	36	splice	287	3	3190
16	brea.cancer	38	2	286	37	vehicle	18	4	846
17	breast-w	9	2	699	38	vote	16	2	435
18	col10	7	10	2019	39	vowel	11	11	990
19	colic	60	2	368	40	waveform	40	3	5000
20	credit-a	42	2	690	41	zoo	16	4	84
21	credit-g	59	2	1000					

Figure B: The decision boundaries of the univariate and improved space trees over a non-linearly separated data set
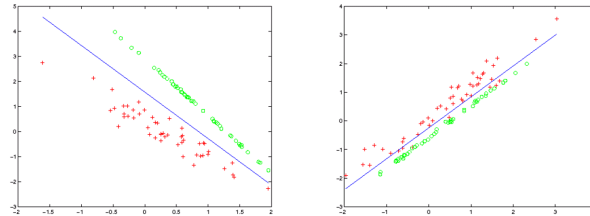

Figure C: Some appropriate class distributions for improved spaces and corresponding W's

Table B: Classification accuracies of Improved and original versions of Bagging

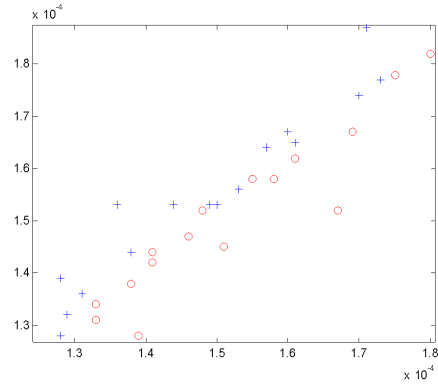| DatasetID	BG	IS-f2-k1-BG	IS-f4-k1-BG	IS-f2-k2-BG	IS-f4-k2-BG
1	53.58	53.52	53.57	54.11	54.07
2	86.66	86.7	87.51	86.7	87.57
3	81.87	82.74	84.23	83.23	84.77
4	84.95	84.98	86.3	84.77	86.48
5	55.83	97.52	98.99	95.41	98.15
6	92.7	91.31	88.52	91.97	91.72
7	65.31	75.69	74.56	76.69	77.67
8	92.29	92.76	93.27	92.81	93.36
9	76.19	76.34	75.87	76.4	76.31
10	92.68	95.42	95.56	95.49	96.27
11	59.23	59.55	60.29	59.84	60.03
12	24.72	25.06	25.53	25.17	25.55
13	86.04	85.62	89.23	85.44	89.41
14	73.22	73.02	75	73.71	74.65
15	83.68	90.48	90.8	89.76	93.47
16	71.68	71.43	70.45	71.15	71.22
17	95.75	96.8	96.68	96.75	96.57
18	78.52	79.06	79.06	79.42	79.1
19	83.1	84.02	84.08	83.83	83.94
20	86.07	86.35	86.61	86.52	86.48
21	74.05	75.17	75.35	75.23	75.56
22	74.88	75.73	75.14	75.98	75.78
23	72.73	70.78	70.54	71.22	68.73
24	80.48	82.15	84.3	81.89	83.11
25	82.71	83.94	84	83.74	84.26
26	91.14	91.85	92.42	92.54	93.7
27	95.4	95.4	95.4	95.53	97
28	88.07	86.32	89.12	85.96	89.3
29	91.65	94.07	94.56	93.51	93.91
30	82.18	81.69	83.24	81.69	84.01
31	47.62	48.97	48.91	49.01	49.17
32	94.83	96.46	96.39	97.94	98.43
33	96.15	97.12	97.44	97.03	97.49
34	76.39	77.31	78.61	79.86	82.93
35	91.3	92.19	91.87	92.34	92
36	94.94	95.18	95.28	95.14	95.27
37	73.22	77.7	77.94	76.95	78.2
38	95.33	95.61	95.75	95.49	95.45
39	82.25	85.79	86.42	85.91	88.15
40	83.34	84.31	85.69	84.34	85.88
41	92.62	97.74	95.6	97.74	95.6

Figure D: Hillvalley data set's class distribution for the 5th and 50th feature pairs

Table C: Classification accuracies of Improved and original versions of Random Forest

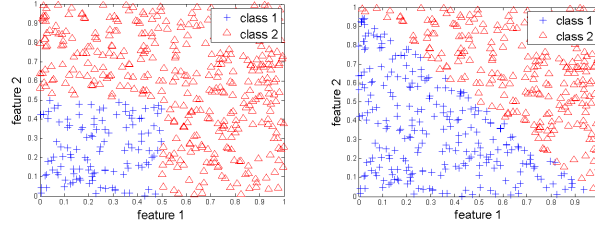| DatasetID	RF	IS-f2-k1-RF	IS-f4-k1-RF	IS-f2-k2-RF	IS-f4-k2-RF
1	54.02	53.87	53.45	53.66	54.15
2	88.88	89.56	89.51	89.93	89.84
3	82.32	83.48	84.65	83.77	84.68
4	84.59	85.99	86.48	86.15	86.88
5	54.6	96.62	99.16	95.61	98.18
6	92.21	91.97	91.39	92.38	91.72
7	67.92	72.61	75.14	74.25	77.44
8	93.29	93.27	93.36	93.27	93.38
9	75.99	75.76	76.43	76.66	76.07
10	95.23	96.08	96.34	97.25	98.1
11	59.74	59.88	60.41	60.22	60.55
12	24.97	25.19	25.31	25.45	25.65
13	85.21	87.75	87.46	86.33	86.04
14	70.99	70.74	70.99	71.14	73.07
15	84.43	90.54	91.1	89.66	92.64
16	71.36	71.4	71.75	71.82	71.47
17	96.15	96.74	96.88	96.67	96.68
18	78.81	79.28	78.85	78.96	79.15
19	83.8	83.99	84.32	84.38	83.86
20	86.7	86.35	86.93	86.86	86.96
21	74.38	74.49	75.92	74.93	75.51
22	75.39	75.96	75.4	76.56	76.5
23	70.59	72.88	70.1	72	71.12
24	80.67	81.96	83.07	82.33	82.93
25	83.81	84.77	82.97	83.68	84
26	92.48	93.25	92.88	92.85	93.68
27	95.27	95.33	95.67	95.27	95.93
28	88.77	91.93	91.05	91.23	91.58
29	93.65	94.42	94.68	94.21	94.43
30	82.18	83.8	83.66	84.01	84.3
31	48.68	50.76	49.54	48.64	49.97
32	94.98	96.5	96.46	98.06	98.45
33	96.82	97.28	97.44	97.19	97.42
34	76.97	78.41	78.17	78.85	81.83
35	91.9	91.66	91.84	91.99	92.56
36	95.96	96.02	95.99	96.02	96.03
37	74.4	76.05	76.83	76.26	77.87
38	95.93	95.93	95.89	95.82	96.07
39	85.07	87.71	86.89	88.03	89.55
40	84.71	85.03	85.85	85.29	86.08
41	98.69	97.5	96.43	96.79	98.69

Figure E: the relationship between the data distribution and tree size. left: univariate (T_u) and bivariate (T_{imp}) trees have the same sizes, right: bivarite tree has less node than the univariate tree.

Table D: Classification accuracies of Improved and original versions of Rotation Forest

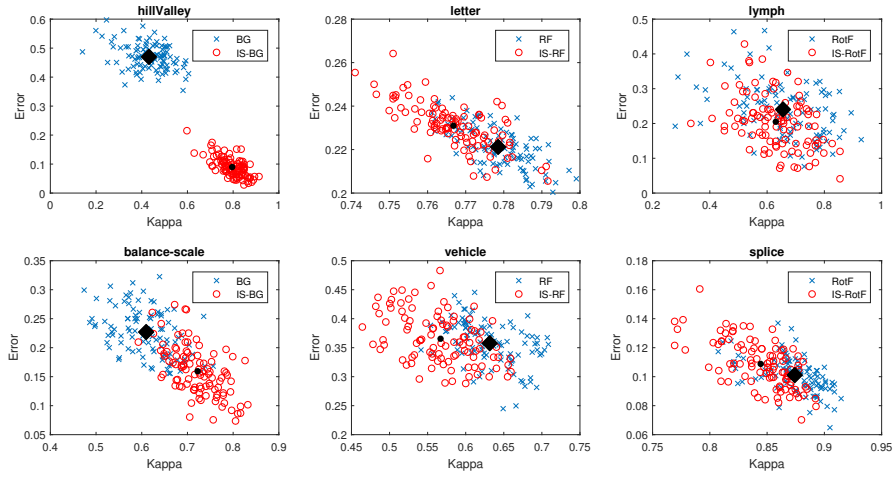| DatasetID	RotF	IS-f2-k1-RotF	IS-f4-k1-RotF	IS-f2-k2-RotF	IS-f4-k2-RotF
1	53.96	53.91	53.25	54.05	53.32
2	87.12	87.61	88.76	87.74	88.74
3	82.68	83.45	82.74	83.94	84.52
4	86.94	87.06	87.46	87.03	87.13
5	98.88	99.08	99.17	98.38	98.14
6	89.92	89.67	88.93	90.66	91.15
7	79.03	78.69	78.61	79.47	80.22
8	93.46	93.43	93.48	93.56	93.55
9	77.22	77.06	76.31	76.8	76.14
10	96.86	96.8	97.91	97.65	98.24
11	60.18	60.45	60.07	60.39	59.89
12	25.53	25.33	25.47	25.27	25.81
13	87.57	88.17	87.99	88.4	87.69
14	73.32	73.17	72.92	73.71	73.76
15	89.06	90.42	92.38	90.06	94.67
16	70.35	71.78	72.24	72.06	72.48
17	96.85	96.88	96.78	97.05	96.58
18	78.56	78.44	77.97	78.78	77.97
19	84.62	83.8	84.46	84.32	84.05
20	87.12	86.74	86.23	86.48	86
21	75.06	75.6	74.79	75.33	74.66
22	76.03	76.58	75.05	77.01	75.7
23	71.85	69.76	68.59	69.37	68.24
24	82.52	82.3	82.52	83.19	82.74
25	83.1	84.65	83.74	84.58	83.87
26	93.82	93.25	93.11	93.05	93.45
27	94.87	96.13	96.27	96.33	97.6
28	87.54	91.58	91.75	91.93	91.93
29	94.63	94.78	94.86	94.57	94.76
30	80.77	84.86	82.89	85.07	83.38
31	48.05	49.37	48.31	49.87	47.95
32	96.78	96.85	96.79	98.46	98.62
33	97.41	97.52	97.54	97.48	97.52
34	80.34	80.72	78.7	80.24	81.35
35	92.5	92.09	92.03	92.21	92.13
36	94.98	95.28	95.59	95.22	95.69
37	77.32	78.43	79.16	77.98	78.92
38	95.93	95.4	95.82	95.33	95.68
39	90.59	89.83	89.59	90.94	91.86
40	85.82	85.74	86.35	85.89	86.38
41	96.9	96.19	93.93	96.07	94.05

Figure F: The Error-Kappa diagrams for Improved Space ($f = 2$ and $k = 1$) and original versions of BG, RF, and $RotF$ over some data sets