

Additional Tables for “A feature construction
framework based on outlier detection and
discriminative pattern mining”

Albrecht Zimmermann

April 17, 2014

Data set	Classes	Attributes	Instances
breast-w	2	10	699
iris	3	5	153
segment	7	20	2310
diabetes	2	9	768
segnoise	7	70	1500
ecoli	8	8	337
liver-disorders	2	7	345
sonar	2	61	208
glass	7	10	214
optdigits	10	65	5619
spambase	2	58	4601
heart-statlog	2	14	270
page-blocks	5	11	5473
spectrometer	48	102	531
ionosphere	2	35	355
pendigits	10	17	10991
vehicle	4	19	849
audiology	24	70	226
molecular-biology promoters	2	58	106
breast-cancer	2	10	286
nursery	5	9	12961
car	4	7	1728
postoperative-patient-data	3	9	90
primary-tumor	22	18	339
dermatology	6	34	365
soybean	19	36	687
kr-vs-kp	2	37	3199
splice	3	61	3189
lung-cancer	3	57	32
tic-tac-toe	2	10	958
lymph	4	19	147
vote	2	17	438

Table 1: Data set characteristics

Data set	Original		CobFC		DC-Fringe		Baseline	
	Train	Test	Train	Test	Train	Test	Train	Test
breast-w	96.11 ± 0.26	96.14 ± 2.24	96.82 ± 0.31	96.43 ± 2.80	97.30 ± 0.45	96.00 ± 2.59	97.98 ± 0.53	95.43 ± 3.21
diabetes	76.27 ± 0.73	76.18 ± 4.83	75.90 ± 1.09	74.62 ± 4.52	82.70 ± 2.89	70.83 ± 3.20	77.39 ± 0.74	74.62 ± 5.92
ecoli	88.33 ± 1.43	85.43 ± 3.82	85.75 ± 1.31	83.65 ± 6.25	88.39 ± 1.97	78.24 ± 7.09	90.68 ± 1.27	84.53 ± 4.76
glass	54.77 ± 2.79	47.36 ± 11.52	76.84 ± 2.81	65.06 ± 12.77	96.88 ± 3.15	72.94 ± 10.37	52.68 ± 7.03	40.28 ± 7.83
heart-statlog	85.76 ± 1.12	83.70 ± 7.45	86.05 ± 0.74	84.07 ± 8.74	92.88 ± 2.89	81.85 ± 6.86	84.93 ± 1.51	83.33 ± 7.04
ionosphere	83.57 ± 0.87	82.63 ± 5.76	89.02 ± 1.94	87.75 ± 3.80	95.70 ± 2.07	89.46 ± 4.67	85.28 ± 1.92	84.90 ± 5.86
iris	95.85 ± 0.86	96.00 ± 4.66	92.52 ± 1.28	91.33 ± 6.32	98.22 ± 1.45	94.67 ± 5.26	96.75 ± 0.70	96.00 ± 6.44
liver-disorders	56.88 ± 2.75	53.34 ± 5.82	70.40 ± 2.08	63.18 ± 6.60	94.82 ± 2.71	66.99 ± 6.50	65.76 ± 3.17	63.53 ± 5.46
optdigits	91.82 ± 0.16	91.30 ± 1.03	92.55 ± 0.38	92.17 ± 1.32	93.93 ± 0.43	93.10 ± 1.25	92.88 ± 0.17	90.87 ± 1.07
pendigits	85.86 ± 0.20	85.72 ± 0.92	84.57 ± 1.08	84.42 ± 1.47	88.42 ± 0.39	88.24 ± 1.08	87.02 ± 0.22	85.32 ± 1.10
spambase	79.61 ± 0.19	81.01 ± 1.01	91.40 ± 0.25	91.57 ± 0.56	93.68 ± 0.56	93.19 ± 1.26	79.57 ± 0.48	78.85 ± 1.80
page-blocks	90.31 ± 0.92	90.28 ± 1.77	90.22 ± 0.39	90.01 ± 1.90	90.48 ± 1.69	90.01 ± 2.55	91.32 ± 5.08	87.96 ± 5.93
segment	80.49 ± 0.24	80.17 ± 2.23	90.70 ± 1.06	89.70 ± 1.82	92.69 ± 1.19	91.69 ± 1.55	84.83 ± 1.21	83.03 ± 1.08
segnoise	81.88 ± 0.30	80.87 ± 1.78	90.31 ± 0.85	89.20 ± 3.20	91.99 ± 2.35	89.20 ± 2.45	83.15 ± 0.76	80.33 ± 3.27
sonar	72.28 ± 1.35	67.83 ± 7.99	72.27 ± 1.71	69.26 ± 6.73	93.70 ± 4.66	77.45 ± 10.97	71.40 ± 1.54	65.95 ± 8.65
spectrometer	62.08 ± 1.08	42.74 ± 5.63	69.35 ± 1.05	43.30 ± 5.91	87.59 ± 1.77	42.55 ± 4.51	69.51 ± 0.94	41.61 ± 7.48
vehicle	46.93 ± 0.76	45.27 ± 2.26	49.72 ± 2.03	49.75 ± 3.86	53.99 ± 2.85	52.01 ± 4.11	52.60 ± 0.67	49.63 ± 3.39
audiology	93.76 ± 0.90	75.24 ± 5.85	91.99 ± 0.52	76.52 ± 7.77	86.82 ± 2.60	74.82 ± 6.06	79.17 ± 1.21	65.08 ± 7.19
breast-cancer	74.98 ± 1.27	72.38 ± 8.80	75.37 ± 1.47	70.18 ± 10.72	81.70 ± 1.31	67.49 ± 10.90	78.44 ± 0.91	72.34 ± 7.57
car	88.22 ± 0.26	86.86 ± 2.64	87.98 ± 2.23	87.03 ± 3.73	91.97 ± 7.41	91.44 ± 8.10	91.18 ± 0.81	81.60 ± 2.77
dermatology	99.33 ± 0.24	98.92 ± 1.40	98.85 ± 0.42	97.82 ± 2.79	99.21 ± 0.43	95.11 ± 2.47	99.25 ± 0.35	98.38 ± 2.61
kr-vs-kp	85.79 ± 0.37	85.39 ± 2.13	87.85 ± 2.44	87.20 ± 2.92	82.50 ± 3.66	81.70 ± 5.30	88.37 ± 0.34	87.58 ± 2.66
lung-cancer	91.31 ± 2.50	52.50 ± 32.41	91.65 ± 2.49	49.17 ± 28.45	97.57 ± 2.33	36.67 ± 23.31	89.13 ± 3.08	51.67 ± 27.72
lymph	88.29 ± 1.11	82.57 ± 9.93	88.14 ± 2.08	83.14 ± 6.38	93.99 ± 1.84	84.48 ± 8.30	89.91 ± 1.49	83.81 ± 9.03
molecular-biology-promoters	98.95 ± 0.49	91.36 ± 7.34	98.95 ± 0.86	91.27 ± 8.51	99.58 ± 0.73	87.64 ± 6.50	100.00 ± 0.00	50.00 ± 3.71
nursery	90.40 ± 0.13	90.31 ± 1.11	93.76 ± 0.40	93.83 ± 0.88	90.49 ± 3.04	90.47 ± 2.94	95.04 ± 0.21	92.52 ± 0.82
postoperative-patient-data	67.04 ± 3.08	54.44 ± 12.23	67.28 ± 5.05	56.67 ± 13.30	79.38 ± 6.32	53.33 ± 17.21	73.17 ± 2.91	58.89 ± 15.76
primary-tumor	56.11 ± 0.83	43.98 ± 5.12	54.70 ± 1.59	42.48 ± 7.61	51.89 ± 2.58	38.95 ± 5.41	56.23 ± 0.89	44.27 ± 7.09
soybean	93.80 ± 0.41	93.26 ± 3.47	94.13 ± 0.49	93.56 ± 2.87	94.29 ± 1.58	91.51 ± 3.23	92.56 ± 0.25	90.19 ± 2.94
splice	96.12 ± 0.19	95.89 ± 1.17	95.09 ± 0.34	94.98 ± 1.17	93.58 ± 1.69	91.99 ± 1.92	96.21 ± 0.15	95.80 ± 1.17
tic-tac-toe	69.38 ± 1.05	67.74 ± 3.90	69.90 ± 0.98	68.48 ± 5.16	91.01 ± 2.21	89.35 ± 2.72	72.90 ± 1.44	68.37 ± 3.94
vote	91.55 ± 0.65	91.50 ± 4.16	94.74 ± 0.67	94.25 ± 3.45	92.26 ± 0.97	90.33 ± 5.06	89.95 ± 0.61	90.58 ± 4.73

Table 2: Full training and testing accuracies for CobFC, DC-Fringe, and a baseline of removing class outliers, Naive Bayes

	Original		CobFC		DC-Fringe		Baseline	
breast-w	97.12 ± 0.23	96.57 ± 1.80	97.33 ± 0.39	96.43 ± 1.81	99.59 ± 0.77	95.57 ± 2.37	98.33 ± 0.72	95.14 ± 1.80
diabetes	77.79 ± 1.29	74.62 ± 6.48	78.86 ± 1.79	74.49 ± 5.99	99.65 ± 1.05	69.14 ± 4.81	79.53 ± 1.00	75.66 ± 5.18
ecoli	89.58 ± 1.54	87.82 ± 5.41	86.97 ± 5.98	84.22 ± 4.30	99.17 ± 0.98	76.18 ± 8.91	91.88 ± 2.07	86.33 ± 5.37
glass	82.29 ± 4.44	70.15 ± 7.41	96.15 ± 6.38	75.30 ± 8.52	99.95 ± 0.16	73.42 ± 8.29	83.24 ± 4.62	68.70 ± 8.95
heart-statlog	86.42 ± 1.21	80.00 ± 8.76	88.64 ± 1.01	82.22 ± 8.69	99.22 ± 2.08	75.56 ± 8.04	85.34 ± 4.50	77.41 ± 4.77
ionosphere	98.26 ± 1.24	93.74 ± 3.75	98.86 ± 1.43	95.16 ± 3.82	99.56 ± 0.50	91.74 ± 4.14	98.17 ± 1.49	92.02 ± 3.51
iris	98.37 ± 1.20	96.67 ± 3.51	96.37 ± 3.21	93.33 ± 6.29	99.48 ± 0.99	95.33 ± 4.50	98.73 ± 0.56	96.00 ± 6.44
liver-disorders	77.46 ± 0.76	73.08 ± 8.01	78.29 ± 2.11	72.24 ± 9.29	99.97 ± 0.10	64.36 ± 7.68	76.36 ± 1.11	69.32 ± 7.73
spambase	95.98 ± 0.82	92.70 ± 1.95	96.31 ± 0.41	94.22 ± 1.09	83.08 ± 40.70	78.52 ± 38.48	95.49 ± 0.69	91.52 ± 2.20
page-blocks	98.21 ± 0.12	96.02 ± 0.68	98.43 ± 0.35	95.89 ± 0.79	88.16 ± 33.06	85.21 ± 31.96	99.20 ± 0.06	94.90 ± 0.83
segment	99.32 ± 0.22	97.36 ± 0.75	99.73 ± 0.16	97.88 ± 1.07	99.97 ± 0.06	97.88 ± 1.07	99.57 ± 0.13	95.67 ± 1.34
segnoise	99.73 ± 0.31	96.33 ± 1.14	99.71 ± 0.33	96.80 ± 1.21	99.73 ± 0.40	97.20 ± 0.98	99.90 ± 0.18	95.40 ± 1.39
sonar	99.73 ± 0.84	86.57 ± 6.25	99.73 ± 0.45	79.76 ± 8.20	100.00 ± 0.00	78.40 ± 9.14	100.00 ± 0.00	86.07 ± 8.28
spectrometer	10.36 ± 0.11	10.36 ± 1.01	10.36 ± 0.11	10.36 ± 1.01	10.36 ± 0.11	10.36 ± 1.01	10.57 ± 0.21	10.36 ± 1.01
vehicle	94.77 ± 1.36	83.45 ± 3.55	95.38 ± 0.95	83.09 ± 2.51	98.00 ± 2.13	80.84 ± 2.83	96.99 ± 2.06	82.14 ± 2.64
audiology	99.80 ± 0.34	78.72 ± 5.98	99.61 ± 1.08	76.94 ± 7.55	99.95 ± 0.16	80.08 ± 9.71	99.62 ± 0.51	77.43 ± 5.62
breast-cancer	82.86 ± 6.14	72.38 ± 7.11	86.95 ± 7.68	72.02 ± 6.39	97.47 ± 0.88	66.10 ± 8.56	84.59 ± 4.26	71.66 ± 5.94
car	100.00 ± 0.00	99.71 ± 0.41	99.86 ± 0.20	99.25 ± 0.72	100.00 ± 0.00	99.83 ± 0.39	100.00 ± 0.00	95.25 ± 1.77
dermatology	99.24 ± 0.43	98.10 ± 2.87	99.70 ± 0.38	97.55 ± 2.69	100.00 ± 0.00	96.19 ± 2.62	99.35 ± 0.46	97.00 ± 2.69
kr-vs-kp	99.98 ± 0.02	99.72 ± 0.31	99.96 ± 0.06	99.47 ± 0.33	99.99 ± 0.02	99.62 ± 0.29	99.98 ± 0.02	99.19 ± 0.49
lung-cancer	76.00 ± 26.65	40.00 ± 19.56	75.65 ± 26.17	36.67 ± 17.21	100.00 ± 0.00	40.00 ± 25.09	78.49 ± 21.19	36.67 ± 17.21
lymph	96.02 ± 4.89	82.57 ± 8.88	96.03 ± 4.26	81.05 ± 8.36	99.62 ± 1.19	79.71 ± 10.97	97.88 ± 2.83	84.52 ± 8.27
molecular-biology.promoters	99.68 ± 0.51	90.36 ± 6.68	99.89 ± 0.33	90.36 ± 7.94	99.79 ± 0.44	88.55 ± 7.56	100.00 ± 0.00	50.00 ± 3.71
postoperative-patient-data	71.11 ± 0.64	71.11 ± 5.74	71.11 ± 0.64	71.11 ± 5.74	88.15 ± 9.44	56.67 ± 22.50	73.28 ± 3.08	67.78 ± 11.05
primary-tumor	63.68 ± 4.50	48.09 ± 6.73	60.54 ± 3.38	49.27 ± 7.88	65.06 ± 6.15	44.29 ± 7.90	62.59 ± 3.15	46.61 ± 7.97
soybean	97.48 ± 1.47	93.12 ± 2.86	97.67 ± 1.09	93.12 ± 3.60	99.67 ± 0.44	91.80 ± 3.93	97.50 ± 1.36	92.53 ± 3.42
splice	98.74 ± 1.24	96.83 ± 0.76	99.06 ± 0.80	96.05 ± 0.89	88.52 ± 33.20	84.88 ± 31.84	99.45 ± 0.67	96.36 ± 0.81
tic-tac-toe	100.00 ± 0.00	100.00 ± 0.00	99.40 ± 0.69	98.12 ± 1.61	99.98 ± 0.05	97.91 ± 1.77	99.76 ± 0.27	97.92 ± 1.55
vote	97.93 ± 0.66	96.55 ± 2.69	97.19 ± 1.10	96.10 ± 3.87	98.85 ± 0.69	94.01 ± 3.30	98.34 ± 0.44	97.25 ± 2.80

Table 3: Full training and testing accuracies for CobFC, DC-Fringe, and a baseline of removing class outliers, SVM

	Original			CobFC			DC-Fringe			Baseline		
breast-w	97.73 ± 0.47	94.57 ± 2.84	97.98 ± 0.44	96.00 ± 2.50	99.73 ± 0.11	94.14 ± 2.89	98.36 ± 0.59	93.99 ± 3.35				
diabetes	83.41 ± 2.04	73.57 ± 4.92	91.13 ± 1.58	75.27 ± 5.50	99.57 ± 0.39	70.96 ± 4.42	83.22 ± 1.66	74.22 ± 3.85				
ecoli	93.09 ± 1.26	81.84 ± 5.38	92.03 ± 1.20	83.65 ± 6.01	97.22 ± 0.70	78.26 ± 6.01	94.43 ± 1.84	82.74 ± 6.25				
glass	93.10 ± 1.19	66.88 ± 11.23	91.74 ± 2.25	67.86 ± 11.93	99.12 ± 0.55	72.01 ± 10.06	93.01 ± 2.10	60.30 ± 10.40				
heart-statlog	92.72 ± 1.53	76.30 ± 5.58	92.43 ± 1.53	81.48 ± 6.76	99.42 ± 0.59	74.81 ± 9.04	93.35 ± 1.04	77.78 ± 6.98				
ionosphere	98.42 ± 0.39	90.02 ± 4.53	97.75 ± 1.03	88.03 ± 2.64	99.62 ± 0.44	90.30 ± 5.09	97.96 ± 0.85	88.33 ± 5.74				
iris	98.00 ± 0.78	94.00 ± 6.63	97.85 ± 0.95	93.33 ± 6.29	99.26 ± 0.49	94.67 ± 4.22	98.50 ± 0.68	92.00 ± 8.78				
liver-disorders	85.57 ± 3.32	64.93 ± 5.96	89.05 ± 2.31	65.83 ± 7.26	99.61 ± 0.25	63.23 ± 9.08	81.98 ± 4.48	64.37 ± 7.55				
optdigits	98.00 ± 0.10	90.39 ± 1.01	97.74 ± 0.20	90.09 ± 1.73	98.74 ± 0.12	92.05 ± 0.99	98.35 ± 0.12	90.23 ± 0.98				
pendigits	99.26 ± 0.06	96.37 ± 0.75	98.99 ± 0.13	96.42 ± 0.69	99.46 ± 0.07	96.96 ± 0.72	99.45 ± 0.08	95.53 ± 0.66				
spambase	97.16 ± 0.23	92.22 ± 1.60	97.11 ± 0.19	92.57 ± 0.56	99.62 ± 0.13	92.54 ± 1.14	97.38 ± 0.50	92.76 ± 1.42				
page-blocks	98.59 ± 0.13	97.02 ± 0.60	98.60 ± 0.24	97.17 ± 0.61	99.31 ± 0.09	96.53 ± 0.91	99.49 ± 0.06	95.91 ± 0.91				
segment	99.14 ± 0.14	96.54 ± 1.40	98.86 ± 0.16	96.41 ± 1.51	99.53 ± 0.08	96.80 ± 1.39	99.45 ± 0.15	95.97 ± 1.29				
segnoise	99.02 ± 0.34	94.87 ± 1.66	98.70 ± 0.26	93.87 ± 2.01	99.46 ± 0.13	94.27 ± 1.76	99.22 ± 0.18	93.80 ± 1.48				
sonar	98.13 ± 0.88	71.17 ± 10.30	97.38 ± 0.99	74.52 ± 7.89	99.52 ± 0.59	78.88 ± 8.56	97.77 ± 1.10	73.52 ± 7.79				
spectrometer	90.56 ± 0.58	48.96 ± 4.04	88.34 ± 0.93	48.79 ± 5.04	93.22 ± 1.10	42.74 ± 3.22	91.46 ± 0.99	48.21 ± 3.51				
vehicle	91.53 ± 3.29	74.60 ± 4.24	94.90 ± 0.77	73.76 ± 3.20	95.13 ± 1.23	68.32 ± 4.21	92.76 ± 3.22	70.09 ± 3.60				
audiology	91.15 ± 1.24	77.02 ± 5.65	91.64 ± 1.40	75.26 ± 5.35	96.41 ± 0.84	77.91 ± 8.37	91.65 ± 0.88	73.93 ± 6.19				
breast-cancer	81.89 ± 1.77	72.09 ± 7.10	82.59 ± 2.46	70.96 ± 5.58	94.68 ± 1.41	63.29 ± 6.73	82.12 ± 2.04	70.63 ± 4.64				
car	98.97 ± 0.21	96.24 ± 1.14	98.90 ± 0.59	96.06 ± 3.02	99.99 ± 0.02	99.71 ± 0.41	99.37 ± 0.22	93.92 ± 1.77				
dermatology	98.03 ± 0.36	96.19 ± 3.89	98.24 ± 0.40	96.19 ± 3.89	99.88 ± 0.16	93.46 ± 3.64	98.61 ± 0.31	95.63 ± 4.16				
kr-vs-kp	99.59 ± 0.11	99.37 ± 0.55	99.57 ± 0.14	99.25 ± 0.49	99.91 ± 0.06	99.47 ± 0.47	99.79 ± 0.07	99.16 ± 0.61				
lung-cancer	90.23 ± 4.98	35.00 ± 29.61	90.23 ± 4.98	35.00 ± 29.61	99.64 ± 1.13	36.67 ± 23.31	92.72 ± 4.90	35.83 ± 28.88				
lymph	94.30 ± 1.42	75.57 ± 13.77	95.35 ± 1.45	76.29 ± 15.46	97.90 ± 0.99	80.29 ± 11.02	95.40 ± 1.03	77.67 ± 13.38				
molecular-biology_promoters	96.96 ± 0.92	74.73 ± 19.04	96.43 ± 1.14	75.64 ± 17.66	99.79 ± 0.44	81.09 ± 7.74	100.00 ± 0.00	50.00 ± 3.71				
nursery	99.83 ± 0.02	99.44 ± 0.25	99.95 ± 0.04	99.84 ± 0.09	100.00 ± 0.00	99.98 ± 0.03	99.86 ± 0.03	96.59 ± 0.80				
postoperative-patient-data	73.21 ± 2.02	68.89 ± 10.21	73.58 ± 2.11	68.89 ± 10.21	84.20 ± 4.68	57.78 ± 22.10	73.16 ± 2.40	66.67 ± 11.71				
primary-tumor	62.41 ± 1.47	40.73 ± 9.09	66.14 ± 1.39	43.07 ± 5.09	75.32 ± 1.41	38.35 ± 7.35	65.00 ± 1.48	36.88 ± 9.95				
soybean	96.88 ± 0.68	92.83 ± 3.62	96.96 ± 0.58	91.80 ± 4.22	99.20 ± 0.48	91.81 ± 3.46	96.01 ± 0.41	89.75 ± 4.68				
splice	98.18 ± 0.11	94.17 ± 1.17	98.25 ± 0.11	94.04 ± 1.22	99.26 ± 0.37	92.48 ± 0.84	98.32 ± 0.15	94.08 ± 0.96				
tic-tac-toe	97.68 ± 0.48	93.21 ± 4.08	97.33 ± 0.70	89.98 ± 4.50	99.72 ± 0.31	97.18 ± 2.03	98.17 ± 0.54	90.71 ± 2.32				
vote	97.24 ± 0.31	95.87 ± 2.36	97.70 ± 0.38	96.32 ± 2.22	98.57 ± 0.50	94.72 ± 4.33	97.62 ± 0.39	94.94 ± 2.82				

Table 4: Full training and testing accuracies for CobFC, DC-Fringe, and a baseline of removing class outliers, J48