Variable Selection for BART: An Application to Gene Regulation
Supplementary Materials

Justin Bleich, Adam Kapelner, Edward I. George and Shane T. Jensen
Department of Statistics
The Wharton School, University of Pennsylvania

March 19, 2014

1 RMSE Results for Gene Regulatory Network Data

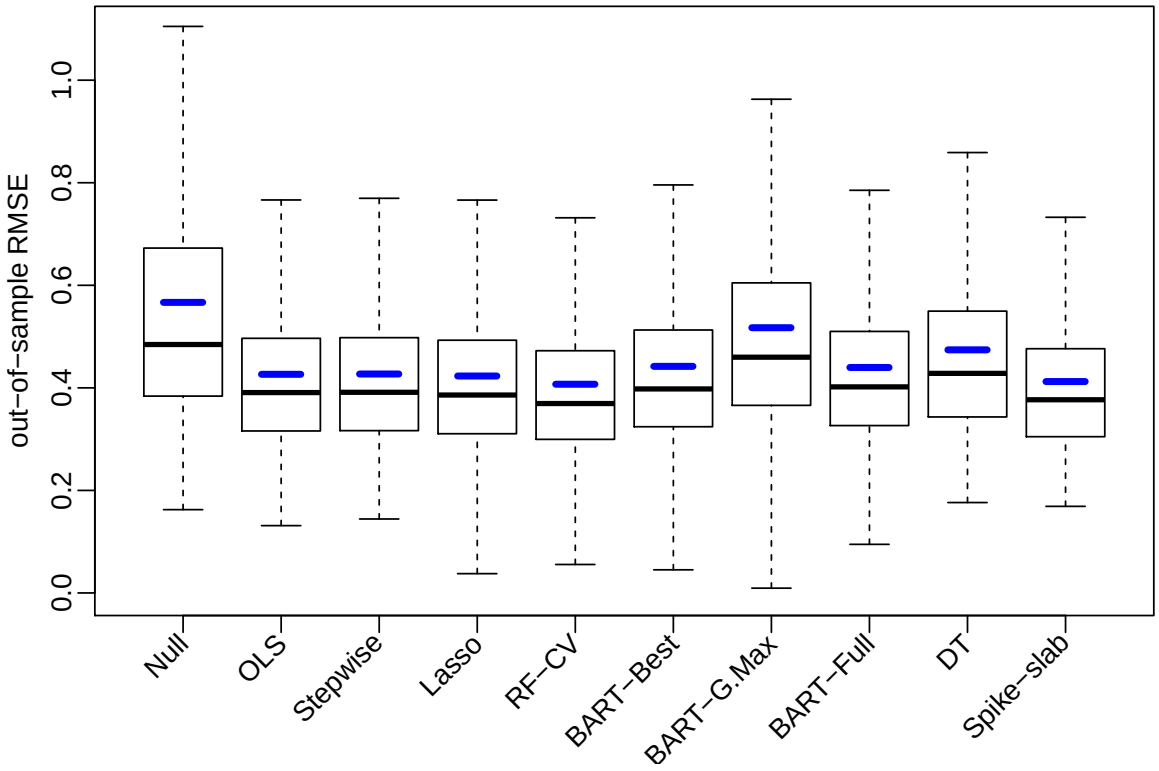

Figure 1: Distributions of out-of-sample RMSEs for each method across the 6026 genes. Blue bars represent the average RMSE across the 6026 genes for each method. Points beyond the whiskers are omitted.

2 Linear Model Simulation Results

	Precision	Recall	F1
BART_CV	0.88	0.84	0.83
BART_pointwise	0.21	1.00	0.34
BART_simul_max	0.97	0.89	0.92
BART_simul_se	0.93	0.90	0.90
BART_CV_good_prior	0.97	0.97	0.96
BART_pointwise_good_prior	0.23	1.00	0.37
BART_simul_max_good_prior	1.00	0.98	0.99
BART_simul_se_good_prior	0.93	1.00	0.96
BART_CV_bad_prior	0.95	0.91	0.92
BART_pointwise_bad_prior	0.23	1.00	0.37
BART_simul_max_bad_prior	0.98	0.84	0.89
BART_simul_se_bad_prior	0.95	0.93	0.93
stepwise_backward	0.02	0.98	0.03
stepwise_forward	0.01	1.00	0.02
lasso	0.86	0.95	0.88
RF_CV	0.40	0.99	0.49
RF_point	0.14	1.00	0.24
RF_simul	0.81	0.99	0.88
dynaTree	0.78	0.65	0.68
spikeSlab	0.51	0.99	0.64

Table 1: Results for $p = 200$, $p_0 = 2$, $\sigma^2 = 5$

	Precision	Recall	F1
BART_CV	0.23	0.25	0.22
BART_pointwise	0.20	0.88	0.32
BART_simul_max	0.20	0.10	0.13
BART_simul_se	0.40	0.30	0.33
BART_CV_good_prior	0.57	0.43	0.45
BART_pointwise_good_prior	0.20	0.97	0.33
BART_simul_max_good_prior	0.52	0.31	0.38
BART_simul_se_good_prior	0.71	0.51	0.57
BART_CV_bad_prior	0.25	0.19	0.20
BART_pointwise_bad_prior	0.19	0.90	0.31
BART_simul_max_bad_prior	0.14	0.07	0.09
BART_simul_se_bad_prior	0.40	0.24	0.29
stepwise_backward	0.02	0.89	0.03
stepwise_forward	0.01	1.00	0.02
lasso	0.22	0.23	0.20
RF_CV	0.13	0.75	0.20
RF_point	0.12	0.73	0.20
RF_simul	0.51	0.47	0.47
dynaTree	0.12	0.11	0.11
spikeSlab	0.54	0.78	0.56

Table 2: Results for $p = 200$, $p_0 = 2$, $\sigma^2 = 20$

	Precision	Recall	F1
BART_CV	0.86	0.86	0.86
BART_pointwise	0.91	0.88	0.89
BART_simul_max	0.26	0.02	0.03
BART_simul_se	0.98	0.22	0.35
BART_CV_good_prior	0.91	0.95	0.93
BART_pointwise_good_prior	0.95	0.95	0.95
BART_simul_max_good_prior	0.76	0.08	0.14
BART_simul_se_good_prior	1.00	0.35	0.51
BART_CV_bad_prior	0.83	0.83	0.83
BART_pointwise_bad_prior	0.88	0.85	0.86
BART_simul_max_bad_prior	0.18	0.02	0.03
BART_simul_se_bad_prior	0.99	0.20	0.32
stepwise_backward	0.16	1.00	0.28
stepwise_forward	0.10	1.00	0.18
lasso	0.54	1.00	0.69
RF_CV	0.55	0.63	0.58
RF_point	0.55	0.66	0.59
RF_simul	0.95	0.30	0.45
dynaTree	0.56	0.07	0.12
spikeSlab	0.55	1.00	0.69

Table 3: Results for $p = 200$, $p_0 = 20$, $\sigma^2 = 5$

	Precision	Recall	F1
BART_CV	0.60	0.41	0.46
BART_pointwise	0.74	0.56	0.63
BART_simul_max	0.20	0.01	0.02
BART_simul_se	0.77	0.08	0.14
BART_CV_good_prior	0.80	0.71	0.75
BART_pointwise_good_prior	0.86	0.74	0.79
BART_simul_max_good_prior	0.40	0.04	0.06
BART_simul_se_good_prior	0.94	0.15	0.26
BART_CV_bad_prior	0.47	0.34	0.37
BART_pointwise_bad_prior	0.69	0.55	0.61
BART_simul_max_bad_prior	0.10	0.00	0.01
BART_simul_se_bad_prior	0.86	0.08	0.14
stepwise_backward	0.14	0.87	0.25
stepwise_forward	0.10	1.00	0.18
lasso	0.68	0.78	0.70
RF_CV	0.44	0.43	0.43
RF_point	0.44	0.42	0.43
RF_simul	0.88	0.14	0.24
dynaTree	0.43	0.05	0.08
spikeSlab	0.58	0.93	0.70

Table 4: Results for $p = 200$, $p_0 = 20$, $\sigma^2 = 20$

	Precision	Recall	F1
BART_CV	0.57	0.82	0.67
BART_pointwise	0.69	0.86	0.77
BART_simul_max	0.18	0.01	0.02
BART_simul_se	0.99	0.26	0.41
BART_CV_good_prior	0.67	0.94	0.78
BART_pointwise_good_prior	0.80	0.95	0.87
BART_simul_max_good_prior	0.54	0.04	0.07
BART_simul_se_good_prior	1.00	0.42	0.58
BART_CV_bad_prior	0.56	0.82	0.67
BART_pointwise_bad_prior	0.67	0.83	0.74
BART_simul_max_bad_prior	0.24	0.01	0.02
BART_simul_se_bad_prior	0.98	0.26	0.41
stepwise_backward			
stepwise_forward			
lasso	0.41	1.00	0.58
RF_CV	0.33	0.48	0.39
RF_point	0.33	0.49	0.39
RF_simul	0.97	0.16	0.27
dynaTree	0.40	0.03	0.05
spikeSlab	0.53	0.96	0.68

Table 5: Results for $p = 500$, $p_0 = 25$, $\sigma^2 = 1$

	Precision	Recall	F1
BART_CV	0.54	0.69	0.58
BART_pointwise	0.60	0.74	0.66
BART_simul_max	0.18	0.01	0.02
BART_simul_se	0.94	0.20	0.32
BART_CV_good_prior	0.63	0.85	0.70
BART_pointwise_good_prior	0.73	0.87	0.79
BART_simul_max_good_prior	0.54	0.03	0.06
BART_simul_se_good_prior	0.99	0.33	0.49
BART_CV_bad_prior	0.54	0.67	0.57
BART_pointwise_bad_prior	0.59	0.73	0.65
BART_simul_max_bad_prior	0.14	0.01	0.01
BART_simul_se_bad_prior	0.97	0.20	0.33
stepwise_backward			
stepwise_forward			
lasso	0.42	1.00	0.59
RF_CV	0.30	0.44	0.35
RF_point	0.28	0.41	0.33
RF_simul	0.92	0.12	0.21
dynaTree	0.28	0.03	0.05
spikeSlab	0.47	0.93	0.62

Table 6: Results for $p = 500$, $p_0 = 25$, $\sigma^2 = 5$

	Precision	Recall	F1
BART_CV	0.53	0.31	0.37
BART_pointwise	0.66	0.43	0.52
BART_simul_max	0.06	0.00	0.00
BART_simul_se	0.95	0.07	0.13
BART_CV_good_prior	0.70	0.57	0.63
BART_pointwise_good_prior	0.79	0.60	0.68
BART_simul_max_good_prior	0.08	0.00	0.00
BART_simul_se_good_prior	0.98	0.12	0.21
BART_CV_bad_prior	0.43	0.26	0.30
BART_pointwise_bad_prior	0.62	0.41	0.49
BART_simul_max_bad_prior	0.04	0.00	0.00
BART_simul_se_bad_prior	0.93	0.07	0.12
stepwise_backward			
stepwise_forward			
lasso	0.38	1.00	0.55
RF_CV	0.34	0.25	0.28
RF_point	0.35	0.25	0.29
RF_simul	0.92	0.04	0.08
dynaTree	0.22	0.01	0.02
spikeSlab	0.64	0.77	0.70

Table 7: Results for $p = 500$, $p_0 = 50$, $\sigma^2 = 1$

	Precision	Recall	F1
BART_CV	0.36	0.18	0.22
BART_pointwise	0.62	0.40	0.48
BART_simul_max	0.00	0.00	0.00
BART_simul_se	0.87	0.05	0.10
BART_CV_good_prior	0.63	0.48	0.54
BART_pointwise_good_prior	0.78	0.58	0.66
BART_simul_max_good_prior	0.12	0.00	0.01
BART_simul_se_good_prior	0.97	0.10	0.17
BART_CV_bad_prior	0.28	0.13	0.15
BART_pointwise_bad_prior	0.56	0.38	0.45
BART_simul_max_bad_prior	0.00	0.00	0.00
BART_simul_se_bad_prior	0.81	0.05	0.09
stepwise_backward			
stepwise_forward			
lasso	0.44	0.97	0.61
RF_CV	0.32	0.21	0.25
RF_point	0.33	0.22	0.26
RF_simul	0.81	0.03	0.06
dynaTree	0.29	0.02	0.03
spikeSlab	0.62	0.74	0.67

Table 8: Results for $p = 500$, $p_0 = 50$, $\sigma^2 = 5$

3 Friedman Simulation Results

	Precision	Recall	F1
BART_CV	0.98	1.00	0.99
BART_pointwise	0.34	1.00	0.50
BART_simul_max	1.00	1.00	1.00
BART_simul_se	1.00	1.00	1.00
BART_CV_good_prior	0.99	1.00	0.99
BART_pointwise_good_prior	0.37	1.00	0.53
BART_simul_max_good_prior	1.00	1.00	1.00
BART_simul_se_good_prior	0.99	1.00	0.99
BART_CV_bad_prior	0.99	1.00	0.99
BART_pointwise_bad_prior	0.33	1.00	0.49
BART_simul_max_bad_prior	1.00	1.00	1.00
BART_simul_se_bad_prior	0.98	1.00	0.99
stepwise_backward			
stepwise_forward			
lasso	0.85	0.80	0.81
RF_CV	0.98	0.84	0.90
RF_point	0.15	0.98	0.26
RF_simul	0.99	0.80	0.88
dynaTree	0.82	0.52	0.62
spikeSlab	0.16	0.81	0.26

Table 9: Results for $p = 500$, $\sigma^2 = 1$

	Precision	Recall	F1
BART_CV	0.83	0.65	0.69
BART_pointwise	0.21	0.93	0.34
BART_simul_max	0.96	0.35	0.50
BART_simul_se	0.91	0.65	0.75
BART_CV_good_prior	0.90	0.73	0.78
BART_pointwise_good_prior	0.23	0.98	0.37
BART_simul_max_good_prior	1.00	0.54	0.69
BART_simul_se_good_prior	0.90	0.74	0.80
BART_CV_bad_prior	0.85	0.71	0.74
BART_pointwise_bad_prior	0.21	0.92	0.34
BART_simul_max_bad_prior	0.96	0.37	0.52
BART_simul_se_bad_prior	0.89	0.64	0.72
stepwise_backward			
stepwise_forward			
lasso	0.85	0.69	0.73
RF_CV	0.63	0.70	0.57
RF_point	0.13	0.79	0.22
RF_simul	0.93	0.62	0.73
dynaTree	0.54	0.20	0.27
spikeSlab	0.11	0.82	0.19

Table 10: Results for $p = 500$, $\sigma^2 = 25$

	Precision	Recall	F1
BART_CV	0.89	1.00	0.93
BART_pointwise	0.13	1.00	0.23
BART_simul_max	1.00	0.99	1.00
BART_simul_se	0.96	1.00	0.97
BART_CV_good_prior	0.95	1.00	0.97
BART_pointwise_good_prior	0.14	1.00	0.25
BART_simul_max_good_prior	1.00	1.00	1.00
BART_simul_se_good_prior	0.97	1.00	0.98
BART_CV_bad_prior	0.91	0.99	0.94
BART_pointwise_bad_prior	0.13	1.00	0.23
BART_simul_max_bad_prior	1.00	1.00	1.00
BART_simul_se_bad_prior	0.95	1.00	0.97
stepwise_backward			
stepwise_forward			
lasso	0.83	0.80	0.80
RF_CV	0.98	0.77	0.86
RF_point	0.09	0.92	0.16
RF_simul	0.98	0.78	0.87
dynaTree	0.79	0.44	0.54
spikeSlab	0.12	0.80	0.21

Table 11: Results for $p = 1000$, $\sigma^2 = 1$

	Precision	Recall	F1
BART_CV	0.70	0.66	0.65
BART_pointwise	0.11	0.88	0.19
BART_simul_max	0.94	0.28	0.42
BART_simul_se	0.79	0.69	0.72
BART_CV_good_prior	0.71	0.71	0.66
BART_pointwise_good_prior	0.12	0.96	0.21
BART_simul_max_good_prior	1.00	0.43	0.59
BART_simul_se_good_prior	0.79	0.75	0.75
BART_CV_bad_prior	0.69	0.69	0.65
BART_pointwise_bad_prior	0.10	0.87	0.18
BART_simul_max_bad_prior	0.98	0.32	0.47
BART_simul_se_bad_prior	0.77	0.69	0.72
stepwise_backward			
stepwise_forward			
lasso	0.77	0.70	0.69
RF_CV	0.47	0.66	0.41
RF_point	0.07	0.73	0.13
RF_simul	0.97	0.57	0.71
dynaTree	0.39	0.17	0.22
spikeSlab	0.09	0.78	0.16

Table 12: Results for $p = 1000$, $\sigma^2 = 25$