

Best Model Selector for Phone Activity Classification

In the following notebook, we are going to be using a dataset from the UCI Machine Learning Repository. The dataset has 561 attributes, which are all gyroscope measurements, and they are all standardized. Our goal is to classify the activity level on a scale from 1 to 6. More information about the dataset:

<https://www.apispreadsheets.com/datasets/122>

Here is some sample data from the dataset:

In [1]:

```
%%bash
head phone_activity.csv
```

```
feature_1,feature_2,feature_3,feature_4,feature_5,feature_6,feature_7,feature_8,feature_9
,feature_10,feature_11,feature_12,feature_13,feature_14,feature_15,feature_16,feature_17,
feature_18,feature_19,feature_20,feature_21,feature_22,feature_23,feature_24,feature_25,f
eature_26,feature_27,feature_28,feature_29,feature_30,feature_31,feature_32,feature_33,fe
ature_34,feature_35,feature_36,feature_37,feature_38,feature_39,feature_40,feature_41,fea
ture_42,feature_43,feature_44,feature_45,feature_46,feature_47,feature_48,feature_49,featu
re_50,feature_51,feature_52,feature_53,feature_54,feature_55,feature_56,feature_57,featu
re_58,feature_59,feature_60,feature_61,feature_62,feature_63,feature_64,feature_65,featur
e_66,feature_67,feature_68,feature_69,feature_70,feature_71,feature_72,feature_73,feature
_74,feature_75,feature_76,feature_77,feature_78,feature_79,feature_80,feature_81,feature_
82,feature_83,feature_84,feature_85,feature_86,feature_87,feature_88,feature_89,feature_9
0,feature_91,feature_92,feature_93,feature_94,feature_95,feature_96,feature_97,feature_98
,feature_99,feature_100,feature_101,feature_102,feature_103,feature_104,feature_105,featu
re_106,feature_107,feature_108,feature_109,feature_110,feature_111,feature_112,feature_11
3,feature_114,feature_115,feature_116,feature_117,feature_118,feature_119,feature_120,fea
ture_121,feature_122,feature_123,feature_124,feature_125,feature_126,feature_127,feature_
128,feature_129,feature_130,feature_131,feature_132,feature_133,feature_134,feature_135,f
eature_136,feature_137,feature_138,feature_139,feature_140,feature_141,feature_142,featur
e_143,feature_144,feature_145,feature_146,feature_147,feature_148,feature_149,feature_150
,feature_151,feature_152,feature_153,feature_154,feature_155,feature_156,feature_157,featu
re_158,feature_159,feature_160,feature_161,feature_162,feature_163,feature_164,feature_1
65,feature_166,feature_167,feature_168,feature_169,feature_170,feature_171,feature_172,fe
ature_173,feature_174,feature_175,feature_176,feature_177,feature_178,feature_179,feature
_180,feature_181,feature_182,feature_183,feature_184,feature_185,feature_186,feature_187,
feature_188,feature_189,feature_190,feature_191,feature_192,feature_193,feature_194,featu
re_195,feature_196,feature_197,feature_198,feature_199,feature_200,feature_201,feature_20
2,feature_203,feature_204,feature_205,feature_206,feature_207,feature_208,feature_209,fea
ture_210,feature_211,feature_212,feature_213,feature_214,feature_215,feature_216,feature_
217,feature_218,feature_219,feature_220,feature_221,feature_222,feature_223,feature_224,f
eature_225,feature_226,feature_227,feature_228,feature_229,feature_230,feature_231,featur
e_232,feature_233,feature_234,feature_235,feature_236,feature_237,feature_238,feature_239
,feature_240,feature_241,feature_242,feature_243,feature_244,feature_245,feature_246,featu
re_247,feature_248,feature_249,feature_250,feature_251,feature_252,feature_253,feature_2
54,feature_255,feature_256,feature_257,feature_258,feature_259,feature_260,feature_261,fe
ature_262,feature_263,feature_264,feature_265,feature_266,feature_267,feature_268,feature
_269,feature_270,feature_271,feature_272,feature_273,feature_274,feature_275,feature_276,
feature_277,feature_278,feature_279,feature_280,feature_281,feature_282,feature_283,featu
re_284,feature_285,feature_286,feature_287,feature_288,feature_289,feature_290,feature_29
1,feature_292,feature_293,feature_294,feature_295,feature_296,feature_297,feature_298,fea
ture_299,feature_300,feature_301,feature_302,feature_303,feature_304,feature_305,feature_
306,feature_307,feature_308,feature_309,feature_310,feature_311,feature_312,feature_313,f
eature_314,feature_315,feature_316,feature_317,feature_318,feature_319,feature_320,featur
e_321,feature_322,feature_323,feature_324,feature_325,feature_326,feature_327,feature_328
,feature_329,feature_330,feature_331,feature_332,feature_333,feature_334,feature_335,featu
re_336,feature_337,feature_338,feature_339,feature_340,feature_341,feature_342,feature_3
43,feature_344,feature_345,feature_346,feature_347,feature_348,feature_349,feature_350,fe
ature_351,feature_352,feature_353,feature_354,feature_355,feature_356,feature_357,feature
_358,feature_359,feature_360,feature_361,feature_362,feature_363,feature_364,feature_365,
feature_366,feature_367,feature_368,feature_369,feature_370,feature_371,feature_372,featu
re_373,feature_374,feature_375,feature_376,feature_377,feature_378,feature_379,feature_38
0,feature_381,feature_382,feature_383,feature_384,feature_385,feature_386,feature_387,fea
ture_388,feature_389,feature_390,feature_391,feature_392,feature_393,feature_394,feature_
395,feature_396,feature_397,feature_398,feature_399,feature_400,feature_401,feature_402,f
```

[illegible]

[illegible]

[illegible]

.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-0.994
,-0.99,-0.992,-0.991,-0.988,-0.994,-1.0,-0.997,-0.907,-1.0,0.394,-0.113,-0.482,-0.996,-0
.994,-0.993,-0.995,-0.982,-0.996,-1.0,-0.994,-1.0,-1.0,0.559,-0.529,-0.859,-0.991,-0.989,
-0.988,-0.991,-0.998,-0.991,-1.0,-0.989,-0.763,-0.897,-0.274,-0.51,-0.831,-0.995,-0.995,
-0.995,-0.996,-0.997,-0.995,-1.0,-0.995,-1.0,-0.937,0.0878,-0.351,-0.699,0.123,0.123,0.69
4,-0.616,-0.848,0.185,-0.0439,5
0.277,-0.0101,-0.105,-0.997,-0.99,-0.995,-0.998,-0.99,-0.996,-0.942,-0.566,-0.823,0.849,0
.696,0.846,-0.994,-1.0,-1.0,-1.0,-0.998,-0.992,-0.995,-0.752,-0.455,-0.551,0.39,-0.182,0.
159,0.187,0.26,-0.243,0.422,-0.418,0.558,-0.218,0.165,0.0809,-0.21,-0.151,0.18,0.968,-0.1
48,0.0919,-0.999,-0.987,-0.997,-0.999,-0.987,-0.997,0.894,-0.167,0.0833,0.988,-0.122,0.09
41,-0.409,0.912,-0.967,-0.985,-0.999,-0.987,-0.998,-1.0,-1.0,-1.0,-0.302,0.379,-0.457,0.5
37,-0.198,0.127,-0.108,0.111,-0.158,0.175,-0.193,0.208,-0.564,0.466,0.443,0.0779,0.0187,0
.00934,-0.995,-0.989,-0.992,-0.995,-0.987,-0.991,-0.994,-0.992,-0.993,0.994,0.993,0.986,-
0.993,-1.0,-1.0,-1.0,-0.995,-0.987,-0.99,-0.779,-0.755,-0.791,0.281,0.1,0.26,0.202,0.351,
-0.0773,0.563,0.352,0.369,0.0434,0.147,0.299,-0.375,0.0418,-0.112,-0.0288,-0.0704,0.079,-
0.985,-0.992,-0.983,-0.986,-0.992,-0.983,-0.87,-0.953,-0.746,0.84,0.911,0.819,-0.985,-1.0
,-1.0,-1.0,-0.985,-0.993,-0.984,-0.364,-0.371,-0.462,-0.403,0.242,0.0591,0.0303,-0.12,0.1
51,-0.381,0.573,0.104,-0.228,0.138,0.346,-0.448,0.645,-0.746,-0.0942,-0.0434,-0.0419,-0.9
92,-0.996,-0.993,-0.993,-0.996,-0.993,-0.98,-0.998,-0.993,0.994,0.997,0.997,-0.995,-1.0,-
1.0,-1.0,-0.994,-0.996,-0.992,-0.628,-0.681,-0.542,-0.265,0.0558,-0.0456,0.236,-0.122,0.2
68,-0.194,-0.00532,0.203,-0.297,0.0699,-0.0123,0.0179,0.187,-0.321,-0.994,-0.995,-0.995,-
0.996,-0.994,-0.994,-1.0,-0.993,-0.785,0.37,-0.393,0.563,-0.482,-0.994,-0.995,-0.995,-0.9
96,-0.994,-0.994,-1.0,-0.993,-0.785,0.37,-0.393,0.563,-0.482,-0.993,-0.995,-0.996,-0.993,
-0.993,-0.993,-1.0,-0.995,-0.871,0.0727,0.0178,-0.0686,-0.166,-0.986,-0.986,-0.986,-0.987
,-0.99,-0.986,-1.0,-0.989,-0.245,-0.173,-0.0078,0.0959,0.0698,-0.995,-0.995,-0.996,-0.992
,-0.993,-0.995,-1.0,-0.996,-0.751,0.179,-0.181,0.0496,-0.169,-0.997,-0.987,-0.993,-0.998,
-0.992,-0.997,-0.997,-0.99,-0.995,-0.999,-0.993,-0.998,-0.996,-0.995,-0.986,-0.995,-1.0,
-1.0,-1.0,-0.995,-0.987,-0.988,-1.0,-0.945,-1.0,-0.613,-1.0,-1.0,0.157,0.319,0.606,-0.631
,-0.935,-0.661,-0.899,-0.881,-0.964,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1
.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1
.0,-1.0,-1.0,-1.0,-1.0,-0.999,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-0.995,-0.988,-0.99
,-0.995,-0.99,-0.993,-0.993,-0.99,-0.992,-0.997,-0.993,-0.994,-0.996,-0.999,-0.985,-0.99
2,-1.0,-1.0,-1.0,-0.991,-0.991,-0.988,-1.0,-1.0,-1.0,-0.32,-0.36,0.52,0.329,0.0548,0.321,
-0.595,-0.905,-0.665,-0.956,-0.723,-0.943,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1
.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1
.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-0.999,-1.0,-1.0,-1.0,-1.0,-0.999,-1.0,-1.0,-0.985,-0.9
93,-0.981,-0.985,-0.992,-0.985,-0.983,-0.993,-0.984,-0.983,-0.993,-0.988,-0.999,-0.998,-
0.982,-0.988,-1.0,-1.0,-1.0,-0.988,-0.994,-0.99,-0.678,-0.808,-0.751,-0.933,-0.935,-0.931
,-0.363,-0.133,0.195,-0.00403,-0.323,-0.29,-0.705,-0.235,-0.608,-1.0,-1.0,-1.0,-1.0,-1.0,
-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0
,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.
0,-1.0,-0.995,-0.995,-0.994,-0.995,-0.989,-0.995,-1.0,-0.992,-1.0,-1.0,0.438,-0.595,-0.80
9,-0.995,-0.995,-0.994,-0.994,-0.983,-0.995,-1.0,-0.997,-1.0,-1.0,0.247,-0.521,-0.803,-0.
991,-0.986,-0.986,-0.988,-0.995,-0.991,-1.0,-0.992,-0.769,-1.0,-0.297,-0.346,-0.727,-0.99
5,-0.995,-0.994,-0.996,-0.999,-0.995,-1.0,-0.995,-1.0,-1.0,0.02,-0.545,-0.845,0.0826,-0.1
43,0.275,-0.368,-0.85,0.185,-0.0421,5
0.279,-0.0196,-0.11,-0.997,-0.967,-0.983,-0.997,-0.966,-0.983,-0.941,-0.566,-0.817,0.851,
0.674,0.834,-0.987,-1.0,-1.0,-1.0,-0.997,-0.972,-0.983,-0.637,-0.515,-0.537,0.36,-0.233,0
.226,0.0695,0.0643,-0.0764,0.138,-0.0368,0.231,-0.115,0.319,-0.488,-0.0959,-0.136,0.611,0
.968,-0.144,0.0931,-0.999,-0.989,-0.992,-0.999,-0.989,-0.992,0.893,-0.166,0.0852,0.988,-0
.12,0.0941,-0.417,0.911,-0.969,-0.985,-0.999,-0.99,-0.991,-1.0,-1.0,-1.0,-0.301,0.401,-0.
503,0.606,-0.423,0.374,-0.362,0.367,-0.128,0.138,-0.147,0.153,0.726,0.939,0.879,0.0822,-0
.017,-0.0158,-0.995,-0.985,-0.988,-0.996,-0.983,-0.989,-0.994,-0.988,-0.982,0.994,0.991,0
.989,-0.992,-1.0,-1.0,-1.0,-0.996,-0.98,-0.99,-0.788,-0.726,-0.749,0.318,0.0311,0.308,0.2
4,0.0785,-0.00397,0.171,0.18,0.293,0.155,0.244,0.612,0.108,0.255,0.265,-0.0286,-0.083,0.0
955,-0.988,-0.989,-0.979,-0.988,-0.99,-0.98,-0.881,-0.954,-0.743,0.839,0.907,0.819,-0.984
,-1.0,-1.0,-1.0,-0.987,-0.992,-0.982,-0.374,-0.57,-0.166,-0.0949,-0.0377,0.335,-0.101,-0.
0469,-0.0549,0.0634,0.24,0.0176,-0.117,0.0488,0.336,-0.0161,0.466,-0.688,-0.0971,-0.0416,
-0.0447,-0.99,-0.995,-0.99,-0.99,-0.996,-0.99,-0.991,-0.996,-0.992,0.992,0.997,0.995,-0.9
94,-1.0,-1.0,-1.0,-0.99,-0.996,-0.989,-0.52,-0.666,-0.5,0.00959,-0.131,0.108,0.417,-0.052
5,0.0512,-0.0398,0.369,0.0208,-0.112,-0.111,0.213,0.218,0.204,-0.416,-0.987,-0.983,-0.987
,-0.978,-0.993,-0.987,-1.0,-0.991,-0.586,0.156,-0.24,0.293,-0.0627,-0.987,-0.983,-0.987,-
0.978,-0.993,-0.987,-1.0,-0.991,-0.586,0.156,-0.24,0.293,-0.0627,-0.991,-0.989,-0.991,-0.
986,-0.993,-0.991,-1.0,-0.993,-0.795,0.178,0.0616,-0.187,-0.207,-0.986,-0.985,-0.982,-0.9
88,-0.993,-0.986,-1.0,-0.98,-0.219,-0.0993,-0.0913,0.196,-0.0113,-0.994,-0.995,-0.995,-0.
996,-0.993,-0.994,-1.0,-0.994,-0.681,0.265,-0.255,0.203,-0.365,-0.996,-0.968,-0.984,-0.99
7,-0.968,-0.983,-0.998,-0.972,-0.983,-0.997,-0.969,-0.977,-0.995,-0.996,-0.996,-0.986,-1
.0,-0.999,-1.0,-0.997,-0.981,-0.978,-1.0,-0.758,-0.869,-0.484,-1.0,-1.0,0.432,0.11,0.38,-
0.321,-0.708,-0.0994,-0.428,-0.263,-0.506,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1.0,-1
.0,-1.0,-1.0,-1.0,-1.0,-0.999,-1.0,-1.0,-0.999,-1.0,-0.999,-0.999,-0.999,-0.999,-1.0,-0.9
99,-0.999,-0.999,-0.999,-1.0,-1.0,-0.999,-1.0,-1.0,-1.0,-0.999,-0.999,-1.0,-1.0,-1.0,-0.9
99,-1.0,-1.0,-0.995,-0.983,-0.987,-0.996,-0.988,-0.988,-0.995,-0.988,-0.988,-0.995,-0.991

[illegible]

Measurements

```
Data Sufficiency: Maybe enough data to generalize. [yellow]
```

Capacity Progression: at [5%, 10%, 20%, 40%, 80%, 100%]
Ideal Machine Learner: 10, 11, 12, 12, 13, 13

Expected Generalization:

Decision Tree: 4.10 bits/bit
Neural Network: 61.08 bits/bit
Random Forest: 96.25 bits/bit

Expected Accuracy:	Training	Validation
Decision Tree:	100.00%	37.25%
Neural Network:	99.65%	98.02%
Random Forest:	100.00%	97.92%

Recommendations:

Time to Build Estimates:

Decision Tree: less than a minute
Neural Network: 31 minutes

End Time: 03/17/2021, 05:19 UTC
Runtime Duration: 3m 51s

For this dataset, we will use a brute force approach for model selection. The tryall.py script allows us to try all 6 possible model configurations on a single dataset in a single line of code. The 3 models are decision tree, neural networks, and random forests, and we can also run each model with the -rank command.

In [3]:

```
python3 tryall.py phone_activity.csv btc
```

```
Error while deleting file : clean005828786797.csv  
Error while deleting file : clean.state  
Cleaning...  
./btc 'phone_activity.csv' -cleanonly --yes  
WARNING: Could not detect a GPU. Neural Network generation will be slow.
```

Brainome Table Compiler 0.991

Copyright (c) 2019-2021 Brainome, Inc. All Rights Reserved.
Licensed to: Alexander Makhratchev (Evaluation)
Expiration Date: 2021-04-30 44 days left
Maximum File Size: 30 GB
Maximum Instances: unlimited
Maximum Attributes: unlimited
Maximum Classes: unlimited
Connected to: daimensions.brainome.ai (local execution)

Command:

```
btc phone_activity.csv -cleanonly --yes
```

Start Time: 03/17/2021, 05:19 UTC

Messages:

Clean only output: clean.csv, clean.state

End Time: 03/17/2021, 05:19 UTC

Runtime Duration: 24s

Done Cleaning!

Splitting Data...

Done Splitting Data!

#####

Running: DT -rank

```
./btc 'train.csv' -headerless -f DT -rank -o DTrank.py -riskoverfit --yes
```

WARNING: Could not detect a GPU. Neural Network generation will be slow.

Brainome Table Compiler 0.991

Copyright (c) 2019-2021 Brainome, Inc. All Rights Reserved.
Licensed to: Alexander Makhratchev (Evaluation)
Expiration Date: 2021-04-30 44 days left
Maximum File Size: 30 GB
Maximum Instances: unlimited
Maximum Attributes: unlimited
Maximum Classes: unlimited
Connected to: daimensions.brainome.ai (local execution)

Command:

btc train.csv -headerless -f DT -rank -o DTrank.py -riskoverfit --yes

Start Time: 03/17/2021, 05:19 UTC

Cleaning...

Attribute Ranking:

Columns selected: 53, 366, 310, 388, 422,
Risk of coincidental column correlation: 0.0%

Test Accuracy Progression:

53 :	56.07%		
366 :	61.27%	change	+5.20%
310 :	66.07%	change	+4.80%
388 :	66.28%	change	+0.21%
422 :	66.32%	change	+0.04%

Pre-training Measurements

Data:

Input:	train.csv (headerless csv)
Target Column:	target
Number of instances:	5149
Number of attributes:	5 out of 561
Number of classes:	6

Class Balance:

0:	16.72%
1:	14.99%
2:	13.65%
3:	17.25%
4:	18.51%
5:	18.88%

Learnability:

Best guess accuracy:	18.88%
Data Sufficiency:	Maybe enough data to generalize. [yellow]

Capacity Progression:

	at [5%, 10%, 20%, 40%, 80%, 100%]
Ideal Machine Learner:	8, 9, 10, 10, 11, 11

Expected Generalization:

Decision Tree:	10.02 bits/bit
Neural Network:	353.00 bits/bit
Random Forest:	13.55 bits/bit

Expected Accuracy:

	Training	Validation
Decision Tree:	89.26%	66.30%
Neural Network:	82.32%	82.10%
Random Forest:	99.22%	84.47%

Recommendations:

Note: Model type DT given by user.

Time to Build Estimates:

DecisionTree:	a few seconds
---------------	---------------

Predictor: DTrank.py
 Classifier Type: Decision Tree
 System Type: 6-way classifier
 Training / Validation Split: Unable to split dataset. The predictor was trained and evaluated on the same data.
 Accuracy:
 Best-guess accuracy: 18.88%

Combined Model Accuracy: 86.50% (4454/5149 correct)

Model Capacity (MEC): 1182 bits

Generalization Ratio: 9.69 bits/bit
 Generalization Index: 3.13
 Percent of Data Memorized: 31.99%
 Resilience to Noise: -0.58 dB

Combined Confusion Matrix:

Actual \ Predicted	0	1	2	3	4	5
0	764	84	11	0	0	2
1	222	533	12	0	0	5
2	68	35	600	0	0	0
3	0	0	0	821	43	24
4	0	0	0	67	880	6
5	17	35	0	40	24	856

Combined Accuracy by Class:

F1	TS	class	TP	FP	TN	FN	TPR	TNR	PPV	NPV
		0	764	307	3981	97	88.73%	97.62%	71.34%	97.62%
09%	65.41%	1	533	154	4223	239	69.04%	94.64%	77.58%	94.64%
06%	57.56%	2	600	23	4423	103	85.35%	97.72%	96.31%	97.72%
50%	82.64%	3	821	107	4154	67	92.45%	98.41%	88.47%	98.41%
42%	82.51%	4	880	67	4129	73	92.34%	98.26%	92.93%	98.26%
63%	86.27%	5	856	37	4140	116	88.07%	97.27%	95.86%	97.27%
80%	84.84%									

End Time: 03/17/2021, 05:23 UTC

Runtime Duration: 3m 26s

Testing on heldout data using DT -rank achieved 62.17% test accuracy

#####

Running: NN -rank

./btc 'train.csv' -headerless -f NN -rank -o NNrank.py -riskoverfit --yes

WARNING: Could not detect a GPU. Neural Network generation will be slow.

Brainome Table Compiler 0.991

Copyright (c) 2019-2021 Brainome, Inc. All Rights Reserved.

Licensed to: Alexander Makhratchev (Evaluation)

Expiration Date: 2021-04-30 44 days left

Maximum File Size: 30 GB

Maximum Instances: unlimited

Maximum Attributes: unlimited

Maximum Classes: unlimited

Connected to: daimensions.brainome.ai (local execution)

Command:

btc train.csv -headerless -f NN -rank -o NNrank.py -riskoverfit --yes

Start Time:

03/17/2021, 05:23 UTC

Attribute Ranking:

Columns selected: 53, 366, 310, 388, 422,
Risk of coincidental column correlation: 0.0%

Test Accuracy Progression:

53	:	56.07%		
366	:	61.27%	change	+5.20%
310	:	66.07%	change	+4.80%
388	:	66.28%	change	+0.21%
422	:	66.32%	change	+0.04%

Pre-training Measurements

Data:

Input:	train.csv (headerless csv)
Target Column:	target
Number of instances:	5149
Number of attributes:	5 out of 561
Number of classes:	6

Class Balance:

0:	16.72%
1:	14.99%
2:	13.65%
3:	17.25%
4:	18.51%
5:	18.88%

Learnability:

Best guess accuracy:	18.88%
Data Sufficiency:	Maybe enough data to generalize. [yellow]

Capacity Progression:

	at [5%, 10%, 20%, 40%, 80%, 100%]
Ideal Machine Learner:	8, 9, 10, 10, 11, 11

Expected Generalization:

Decision Tree:	10.02 bits/bit
Neural Network:	353.00 bits/bit
Random Forest:	13.55 bits/bit

Expected Accuracy:

	Training	Validation
Decision Tree:	89.26%	66.30%
Neural Network:	82.32%	82.10%
Random Forest:	99.22%	84.47%

Recommendations:

Note: Model type NN given by user.

Time to Build Estimates:

Neural Network:	21 minutes
-----------------	------------

Predictor:

Classifier Type:	NNrank.py
System Type:	Neural Network
Training / Validation Split:	6-way classifier
Accuracy:	Unable to split dataset. The predictor was trained and evaluated on the same data.

Best-guess accuracy: 18.88%

Combined Model Accuracy: 85.29% (4392/5149 correct)

Model Capacity (MEC): 132 bits
Model Capacity Utilized: 132 bits
Generalization Ratio: 85.70 bits/bit
Generalization Index: 27.66
Percent of Data Memorized: 3.62%
Resilience to Noise: -1.52 dB

Combined Confusion Matrix:

Actual \ Predicted	0	1	2	3	4	5
0	749	65	47	0	0	0
1	89	656	27	0	0	0
2	72	45	586	0	0	0
3	0	0	0	702	98	88
4	0	0	0	134	807	12
5	0	0	0	53	27	892

Combined Accuracy by Class:

	class	TP	FP	TN	FN	TPR	TNR	PPV	NPV
F1	TS								
58%	73.29%	0	749	161	4127	112	86.99%	97.36%	82.31%
31%	74.38%	1	656	110	4267	116	84.97%	97.35%	85.64%
99%	75.42%	2	586	74	4372	117	83.36%	97.39%	88.79%
01%	65.30%	3	702	187	4074	186	79.05%	95.63%	78.97%
62%	74.86%	4	807	125	4071	146	84.68%	96.54%	86.59%
84%	83.21%	5	892	100	4077	80	91.77%	98.08%	89.92%

End Time: 03/17/2021, 05:30 UTC

Runtime Duration: 7m 37s

Testing on heldout data using NN -rank achieved 83.67% test accuracy

#####

Running: RF -rank

./btc 'train.csv' -headerless -f RF -rank -o RFrank.py -riskoverfit --yes

WARNING: Could not detect a GPU. Neural Network generation will be slow.

Brainome Table Compiler 0.991

Copyright (c) 2019-2021 Brainome, Inc. All Rights Reserved.

Licensed to: Alexander Makhratchev (Evaluation)

Expiration Date: 2021-04-30 44 days left

Maximum File Size: 30 GB

Maximum Instances: unlimited

Maximum Attributes: unlimited

Maximum Classes: unlimited

Connected to: daimensions.brainome.ai (local execution)

Command:

btc train.csv -headerless -f RF -rank -o RFrank.py -riskoverfit --yes

Start Time: 03/17/2021, 05:31 UTC

Attribute Ranking:

Columns selected: 53, 366, 310, 388, 422,

Risk of coincidental column correlation: 0.0%

Test Accuracy Progression:

53	:	56.07%		
366	:	61.27%	change	+5.20%
310	:	66.07%	change	+4.80%
388	:	66.28%	change	+0.21%
422	:	66.32%	change	+0.04%

Pre-training Measurements

Data:

Input:	train.csv (headerless csv)
Target Column:	target
Number of instances:	5149
Number of attributes:	5 out of 561
Number of classes:	6

Class Balance:

0:	16.72%
1:	14.99%
2:	13.65%
3:	17.25%
4:	18.51%
5:	18.88%

Learnability:

Best guess accuracy:	18.88%
Data Sufficiency:	Maybe enough data to generalize. [yellow]

Capacity Progression:

Ideal Machine Learner:	at [5%, 10%, 20%, 40%, 80%, 100%]
------------------------	-------------------------------------

	8, 9, 10, 10, 11, 11
--	----------------------

Expected Generalization:

Decision Tree:	10.02 bits/bit
Neural Network:	353.00 bits/bit
Random Forest:	13.55 bits/bit

Expected Accuracy:

	Training	Validation
Decision Tree:	89.26%	66.30%
Neural Network:	82.32%	82.10%
Random Forest:	99.22%	84.47%

Recommendations:

Note: Model type RF given by user.

Predictor:

Classifier Type:	RFrank.py
System Type:	Random Forest
System Type:	6-way classifier
Training / Validation Split:	Unable to split dataset. The predictor was trained and evaluated on the same data.
Accuracy:	
Best-guess accuracy:	18.88%

Combined Model Accuracy:	98.64% (5079/5149 correct)
--------------------------	----------------------------

Model Capacity (MEC):	13 bits
-----------------------	---------

Generalization Ratio:	1006.39 bits/bit
Generalization Index:	324.77
Percent of Data Memorized:	0.31%
Resilience to Noise:	-2.59 dB

Combined Confusion Matrix:

Actual	Predicted						
0	861	0	0	0	0	0	0
1	0	772	0	0	0	0	0
2	0	0	703	0	0	0	0
3	0	0	0	850	34	4	
4	0	0	0	9	944	0	
5	0	0	0	17	6	949	

Combined Accuracy by Class:

		class	TP	FP	TN	FN	TPR	TNR	PPV	NPV	
F1	TS	0	861	0	4288	0	100.00%	100.00%	100.00%	100.00%	100.
00%	100.00%	1	772	0	4377	0	100.00%	100.00%	100.00%	100.00%	100.
00%	100.00%	2	703	0	4446	0	100.00%	100.00%	100.00%	100.00%	100.
00%	100.00%	3	850	26	4235	38	95.72%	99.11%	97.03%	99.11%	96.
37%	93.00%	4	944	40	4156	9	99.06%	99.78%	95.93%	99.78%	97.
47%	95.07%	5	949	4	4173	23	97.63%	99.45%	99.58%	99.45%	98.
60%	97.23%										

Attribute Ranking:

1 : 42.45%
4 : 22.34%
0 : 18.81%
2 : 12.90%
3 : 3.49%

End Time: 03/17/2021, 05:35 UTC
Runtime Duration: 4m 8s
Testing on heldout data using RF -rank achieved 85.22% test accuracy

Running: RF
./btc 'train.csv' -headerless -f RF -o RF.py -riskoverfit --yes
WARNING: Could not detect a GPU. Neural Network generation will be slow.

Brainome Table Compiler 0.991

Copyright (c) 2019-2021 Brainome, Inc. All Rights Reserved.
Licensed to: Alexander Makhratchev (Evaluation)
Expiration Date: 2021-04-30 44 days left
Maximum File Size: 30 GB
Maximum Instances: unlimited
Maximum Attributes: unlimited
Maximum Classes: unlimited
Connected to: daimensions.brainome.ai (local execution)

Command:

btc train.csv -headerless -f RF -o RF.py -riskoverfit --yes

Start Time: 03/17/2021, 05:35 UTC

Pre-training Measurements

Data:
Input: train.csv (headerless csv)
Target Column: target
Number of instances: 5149
Number of attributes: 561 out of 561
Number of classes: 6

Class Balance:

0: 16.72%
1: 14.99%
2: 13.65%
3: 17.25%
4: 18.51%
5: 18.88%

Learnability:

Best guess accuracy: 18.88%
Data Sufficiency: Maybe enough data to generalize. [yellow]

Capacity Progression: at [5%, 10%, 20%, 40%, 80%, 100%]
Ideal Machine Learner: 9, 10, 11, 11, 12, 12

Expected Generalization:

Decision Tree: 4.16 bits/bit
Neural Network: 41.93 bits/bit
Random Forest: 63.57 bits/bit

Expected Accuracy:	Training	Validation
Decision Tree:	100.00%	38.10%
Neural Network:	99.42%	97.05%
Random Forest:	100.00%	96.86%

Recommendations:

Note: Model type RF given by user.

Predictor:

RF.py
Classifier Type: Random Forest
System Type: 6-way classifier
Training / Validation Split: Unable to split dataset. The predictor was trained and evaluated on the same data.
Accuracy:
Best-guess accuracy: 18.88%

Combined Model Accuracy: 100.00% (5149/5149 correct)

Model Capacity (MEC): 5 bits

Generalization Ratio: 2652.69 bits/bit
Generalization Index: 856.06
Percent of Data Memorized: 0.12%
Resilience to Noise: -3.01 dB

Combined Confusion Matrix:

Actual \ Predicted	0	1	2	3	4	5
0	861	0	0	0	0	0
1	0	772	0	0	0	0
2	0	0	703	0	0	0
3	0	0	0	888	0	0
4	0	0	0	0	953	0
5	0	0	0	0	0	972

Combined Accuracy by Class:

F1	TS	class	TP	FP	TN	FN	TPR	TNR	PPV	NPV
00%	100.00%	0	861	0	4288	0	100.00%	100.00%	100.00%	100.00%
00%	100.00%	1	772	0	4377	0	100.00%	100.00%	100.00%	100.00%
00%	100.00%	2	703	0	4446	0	100.00%	100.00%	100.00%	100.00%
00%	100.00%	3	888	0	4261	0	100.00%	100.00%	100.00%	100.00%

00%	100.00%										
		4	953	0	4196	0	100.00%	100.00%	100.00%	100.00%	100.
00%	100.00%										
		5	972	0	4177	0	100.00%	100.00%	100.00%	100.00%	100.
00%	100.00%										

Attribute Ranking:

348	:	18.90%
503	:	9.82%
296	:	7.39%
504	:	5.94%
330	:	5.57%
502	:	4.94%
52	:	4.94%
16	:	3.27%
40	:	2.97%
129	:	2.36%
65	:	2.35%
50	:	2.14%
352	:	1.78%
7	:	1.34%
474	:	1.33%
450	:	1.30%
426	:	1.21%
69	:	1.17%
537	:	1.06%
316	:	1.01%
166	:	0.90%
445	:	0.65%
302	:	0.62%
94	:	0.62%
38	:	0.62%
49	:	0.58%
409	:	0.56%
371	:	0.54%
423	:	0.52%
245	:	0.48%
464	:	0.48%
418	:	0.42%
200	:	0.42%
138	:	0.41%
104	:	0.37%
75	:	0.37%
174	:	0.37%
9	:	0.35%
454	:	0.32%
518	:	0.30%
54	:	0.29%
310	:	0.28%
63	:	0.26%
300	:	0.25%
448	:	0.24%
113	:	0.23%
74	:	0.22%
101	:	0.19%
89	:	0.19%
446	:	0.19%
132	:	0.18%
428	:	0.17%
179	:	0.17%
434	:	0.17%
278	:	0.17%
274	:	0.16%
106	:	0.16%
37	:	0.16%
209	:	0.14%
57	:	0.14%
451	:	0.14%
139	:	0.13%
317	:	0.13%
201	:	0.13%
22	:	0.12%

369	:	0.12%
419	:	0.12%
432	:	0.12%
142	:	0.11%
42	:	0.10%
427	:	0.10%
61	:	0.09%
429	:	0.09%
557	:	0.09%
196	:	0.09%
41	:	0.08%
540	:	0.08%
53	:	0.08%
449	:	0.08%
510	:	0.08%
55	:	0.08%
458	:	0.08%
207	:	0.07%
159	:	0.07%
249	:	0.07%
45	:	0.07%
141	:	0.07%
527	:	0.07%
184	:	0.07%
558	:	0.06%
28	:	0.06%
185	:	0.06%
460	:	0.06%
370	:	0.06%
459	:	0.06%
468	:	0.06%
374	:	0.06%
233	:	0.06%
186	:	0.05%
58	:	0.05%
433	:	0.05%
298	:	0.05%
62	:	0.05%
556	:	0.04%
390	:	0.04%
351	:	0.04%
297	:	0.04%
243	:	0.04%
265	:	0.04%
78	:	0.04%
136	:	0.04%
198	:	0.04%
373	:	0.04%
73	:	0.03%
322	:	0.03%
102	:	0.03%
127	:	0.03%
487	:	0.03%
199	:	0.03%
381	:	0.03%
112	:	0.03%
270	:	0.03%
172	:	0.03%
133	:	0.03%
36	:	0.03%
46	:	0.03%
291	:	0.03%
160	:	0.03%
23	:	0.03%
309	:	0.02%
126	:	0.02%
148	:	0.02%
331	:	0.02%
301	:	0.02%
248	:	0.02%
461	:	0.02%
121	:	0.02%

119	:	0.02%
551	:	0.02%
150	:	0.02%
187	:	0.02%
204	:	0.02%
197	:	0.02%
285	:	0.02%
295	:	0.02%
140	:	0.02%
385	:	0.02%
505	:	0.02%
202	:	0.02%
143	:	0.02%
91	:	0.02%
275	:	0.02%
162	:	0.02%
247	:	0.02%
80	:	0.01%
383	:	0.01%
115	:	0.01%
194	:	0.01%
24	:	0.01%
149	:	0.01%
456	:	0.01%
511	:	0.01%
117	:	0.01%
299	:	0.01%
235	:	0.01%
114	:	0.01%
268	:	0.01%
267	:	0.01%
118	:	0.01%
77	:	0.01%
329	:	0.01%
457	:	0.01%
92	:	0.01%
372	:	0.01%
286	:	0.01%
158	:	0.01%
43	:	0.01%
103	:	0.01%
1	:	0.01%
271	:	0.01%
358	:	0.01%
51	:	0.01%
228	:	0.01%
190	:	0.01%
35	:	0.01%
526	:	0.01%
236	:	0.01%
161	:	0.01%
79	:	0.01%
354	:	0.01%
145	:	0.00%
122	:	0.00%
416	:	0.00%
116	:	0.00%
13	:	0.00%
413	:	0.00%
56	:	0.00%
212	:	0.00%
134	:	0.00%
157	:	0.00%
163	:	0.00%
528	:	0.00%
108	:	0.00%
130	:	0.00%
180	:	0.00%
39	:	0.00%
363	:	0.00%
76	:	0.00%
250	:	0.00%

379	:	0.00%
6	:	0.00%
365	:	0.00%
559	:	0.00%
171	:	0.00%
466	:	0.00%
110	:	0.00%
85	:	0.00%
12	:	0.00%
273	:	0.00%
120	:	0.00%
152	:	0.00%
64	:	0.00%
519	:	0.00%
71	:	0.00%
67	:	0.00%
555	:	0.00%
380	:	0.00%
560	:	0.00%
72	:	0.00%
128	:	0.00%
230	:	0.00%
251	:	0.00%
444	:	0.00%
318	:	0.00%
455	:	0.00%
25	:	0.00%
453	:	0.00%
0	:	0.00%
105	:	0.00%
261	:	0.00%
264	:	0.00%
338	:	0.00%
452	:	0.00%
192	:	0.00%
525	:	0.00%
59	:	0.00%
479	:	0.00%
182	:	0.00%
375	:	0.00%
495	:	0.00%
544	:	0.00%
263	:	0.00%
538	:	0.00%
524	:	0.00%
68	:	0.00%
539	:	0.00%
412	:	0.00%
156	:	0.00%
246	:	0.00%
550	:	0.00%
506	:	0.00%
195	:	0.00%
279	:	0.00%
447	:	0.00%
294	:	0.00%
2	:	0.00%
553	:	0.00%
144	:	0.00%
32	:	0.00%
210	:	0.00%
473	:	0.00%
155	:	0.00%
66	:	0.00%
238	:	0.00%
167	:	0.00%
292	:	0.00%
84	:	0.00%

End Time: 03/17/2021, 05:37 UTC
Runtime Duration: 2m 30s
Testing on heldout data using RF achieved 98.54% test accuracy

Running: NN
./btc 'train.csv' -headerless -f NN -o NN.py -riskoverfit --yes
WARNING: Could not detect a GPU. Neural Network generation will be slow.

Brainome Table Compiler 0.991

Copyright (c) 2019-2021 Brainome, Inc. All Rights Reserved.
Licensed to: Alexander Makhratchev (Evaluation)
Expiration Date: 2021-04-30 44 days left
Maximum File Size: 30 GB
Maximum Instances: unlimited
Maximum Attributes: unlimited
Maximum Classes: unlimited
Connected to: daimensions.brainome.ai (local execution)

Command:

btc train.csv -headerless -f NN -o NN.py -riskoverfit --yes

Start Time: 03/17/2021, 05:37 UTC

Pre-training Measurements

Data:
Input: train.csv (headerless csv)
Target Column: target
Number of instances: 5149
Number of attributes: 561 out of 561
Number of classes: 6

Class Balance:
0: 16.72%
1: 14.99%
2: 13.65%
3: 17.25%
4: 18.51%
5: 18.88%

Learnability:
Best guess accuracy: 18.88%
Data Sufficiency: Maybe enough data to generalize. [yellow]

Capacity Progression: at [5%, 10%, 20%, 40%, 80%, 100%]
Ideal Machine Learner: 9, 10, 11, 11, 12, 12

Expected Generalization:
Decision Tree: 4.16 bits/bit
Neural Network: 41.93 bits/bit
Random Forest: 63.57 bits/bit

Expected Accuracy:	Training	Validation
Decision Tree:	100.00%	38.10%
Neural Network:	99.42%	97.05%
Random Forest:	100.00%	96.86%

Recommendations:
Note: Model type NN given by user.

Time to Build Estimates:
Neural Network: 35 minutes

Predictor: NN.py
Classifier Type: Neural Network
System Type: 6-way classifier
Training / Validation Split: Unable to split dataset. The predictor was trained and evaluated on the same data.
Accuracy:
Best-guess accuracy: 18.88%

Combined Model Accuracy: 99.96% (5147/5149 correct)

Model Capacity (MEC): 3414 bits
Model Capacity Utilized: 3414 bits
Generalization Ratio: 3.86 bits/bit
Generalization Index: 1.25
Percent of Data Memorized: 80.20%
Resilience to Noise: -0.18 dB

Combined Confusion Matrix:

Actual \ Predicted	0	1	2	3	4	5
0	861	0	0	0	0	0
1	0	772	0	0	0	0
2	0	0	703	0	0	0
3	0	0	0	887	1	0
4	0	0	0	1	952	0
5	0	0	0	0	0	972

Combined Accuracy by Class:

	class	TP	FP	TN	FN	TPR	TNR	PPV	NPV
F1	TS								
00%	100.00%	0 861	0	4288	0	100.00%	100.00%	100.00%	100.00%
00%	100.00%	1 772	0	4377	0	100.00%	100.00%	100.00%	100.00%
00%	100.00%	2 703	0	4446	0	100.00%	100.00%	100.00%	100.00%
89%	99.78%	3 887	1	4260	1	99.89%	99.98%	99.89%	99.98%
90%	99.79%	4 952	1	4195	1	99.90%	99.98%	99.90%	99.98%
00%	100.00%	5 972	0	4177	0	100.00%	100.00%	100.00%	100.00%

End Time: 03/17/2021, 05:51 UTC

Runtime Duration: 13m 59s

Testing on heldout data using NN achieved 97.88% test accuracy

#####

Running: DT

./btc 'train.csv' -headerless -f DT -o DT.py -riskoverfit --yes

WARNING: Could not detect a GPU. Neural Network generation will be slow.

Brainome Table Compiler 0.991

Copyright (c) 2019-2021 Brainome, Inc. All Rights Reserved.

Licensed to: Alexander Makhratchev (Evaluation)

Expiration Date: 2021-04-30 44 days left

Maximum File Size: 30 GB

Maximum Instances: unlimited

Maximum Attributes: unlimited

Maximum Classes: unlimited

Connected to: daimensions.brainome.ai (local execution)

Command:

btc train.csv -headerless -f DT -o DT.py -riskoverfit --yes

Start Time: 03/17/2021, 05:52 UTC

Pre-training Measurements

Data:
Input: train.csv (headerless csv)
Target Column: target
Number of instances: 5149
Number of attributes: 561 out of 561
Number of classes: 6

Class Balance:
0: 16.72%
1: 14.99%
2: 13.65%
3: 17.25%
4: 18.51%
5: 18.88%

Learnability:
Best guess accuracy: 18.88%
Data Sufficiency: Maybe enough data to generalize. [yellow]

Capacity Progression: at [5%, 10%, 20%, 40%, 80%, 100%]
Ideal Machine Learner: 9, 10, 11, 11, 12, 12

Expected Generalization:
Decision Tree: 4.16 bits/bit
Neural Network: 41.93 bits/bit
Random Forest: 63.57 bits/bit

Expected Accuracy:	Training	Validation
Decision Tree:	100.00%	38.10%
Neural Network:	99.42%	97.05%
Random Forest:	100.00%	96.86%

Recommendations:
Note: Model type DT given by user.

Time to Build Estimates:
Decision Tree: a few seconds

Predictor: DT.py
Classifier Type: Decision Tree
System Type: 6-way classifier
Training / Validation Split: Unable to split dataset. The predictor was trained and evaluated on the same data.
Accuracy:
Best-guess accuracy: 18.88%

Combined Model Accuracy: 99.98% (5148/5149 correct)

Model Capacity (MEC): 3187 bits

Generalization Ratio: 4.15 bits/bit
Generalization Index: 1.34
Percent of Data Memorized: 74.72%
Resilience to Noise: -0.21 dB

Combined Confusion Matrix:

Actual \ Predicted	0	1	2	3	4	5
0	861	0	0	0	0	0
1	1	771	0	0	0	0
2	0	0	703	0	0	0

3	0	0	0	888	0	0
4	0	0	0	0	953	0
5	0	0	0	0	0	972

Combined Accuracy by Class:											
		class	TP	FP	TN	FN	TPR	TNR	PPV	NPV	
F1	TS										
		0	861	1	4287	0	100.00%	100.00%	99.88%	100.00%	99.
94%	99.88%										
		1	771	0	4377	1	99.87%	99.98%	100.00%	99.98%	99.
94%	99.87%										
		2	703	0	4446	0	100.00%	100.00%	100.00%	100.00%	100.
00%	100.00%										
		3	888	0	4261	0	100.00%	100.00%	100.00%	100.00%	100.
00%	100.00%										
		4	953	0	4196	0	100.00%	100.00%	100.00%	100.00%	100.
00%	100.00%										
		5	972	0	4177	0	100.00%	100.00%	100.00%	100.00%	100.
00%	100.00%										

```

End Time:          03/17/2021, 05:53 UTC
Runtime Duration:   1m 43s
Testing on heldout data using DT achieved 37.36% test accuracy
#####
Done Running!
Summary:
{'DT -rank': 62.17, 'NN -rank': 83.67, 'RF -rank': 85.22, 'RF': 98.54, 'NN': 97.88, 'DT': 37.36}
Best Test Accuracy: 98.54
Using: RF
Total Time Elapsed: 2068 seconds
Error while deleting file : clean051920316141.csv
Error while deleting file : clean.state

```

Accuracy on Validation for Each Model:

Best Guess: 22.82%

DT -rank: 62.17%

NN -rank: 83.67%

RF -rank: 85.17%

DT: 37.36%

NN: 97.88%

RF: 98.45%

The random forest model worked the best with 98.45% accuracy, with the neural network followed closely behind. However the tryall script does take a while to run, especially on larger datasets.

Citation:

UCI Machine Learning Repository Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. A Public Domain Dataset for Human Activity Recognition Using Smartphones. 21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2013. Bruges, Belgium 24-26 April 2013. Jorge L. Reyes-Ortiz(1,2), Davide Anguita(1), Alessandro Ghio(1), Luca Oneto(1) and Xavier Parra(2) 1 - Smartlab - Non-Linear Complex Systems Laboratory DITEN - Università degli Studi di Genova, Genoa (I-16145), Italy. 2 - CETpD - Technical Research Centre for Dependency Care and Autonomous Living Universitat Politècnica de Catalunya (BarcelonaTech) Vilanova i la Geltrú (08800) Spain

In []: