Assignment 2 - Randomization Algorithms

Ronny Pena / pena1164@gatech.edu

Dataset from assignment 1: http://archive.ics.uci.edu/ml/datasets/Congressional+Voting+Records

Predicting congressmen party affiliations based on voting behavior

Algorithms evaluated:

- Neural Networks with Randomized Hill Climbing Algorithm
- Neural Networks with Simulated Annealing Algorithm
- Neural Networks with Genetic Algorithm
- Randomized Hill Climbing Algorithm
- Simulated Annealing Algorithm
- Genetic Algorithm
- MIMIC Optimization Algorithm

Data setup:

435 Records, 66% training data / 33% test data

Data conversion:

Attributes/Features: y = 1, n = 0, "?" = -1

Classes: democrat = 0, republican = 1

Data samples:

republican,n,n,n,y,y,y,n,n,n,y,n,y,y,y,n,y

republican,y,y,y,y,y,y,y,y,n,y,?,?,?,y,n,y

democrat,y,y,y,n,n,n,y,y,y,n,n,n,n,n,n,y

democrat,n,y,y,n,n,y,y,y,?,y,n,n,n,n,n,y

republican,y,y,n,y,y,y,n,n,n,y,n,n,y,y,n,y

democrat,n,y,y,n,y,y,y,y,n,n,y,n,y,n,y,y

democrat,n,y,y,n,n,?,y,y,y,y,n,?,y,y,y

Processed data:

[0.0, 1.0, 0.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 1.0, -1.0, 1.0, 1.0, 1.0, 0.0, 1.0] - [1.0]

[0.0, 1.0, 0.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 1.0, 1.0, 0.0, -1.0] - [1.0]

[-1.0, 1.0, 1.0, -1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 1.0, 1.0, 0.0, 0.0] - [0.0]

[1.0, 1.0, 1.0, 0.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, 1.0, -1.0, 1.0, 1.0, 1.0, 1.0] - [0.0]

[0.0, 1.0, 1.0, 0.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 1.0, 1.0, 1.0] - [0.0]

[0.0, 1.0, 0.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, -1.0, 1.0, 1.0, 1.0] - [0.0]

[0.0, 1.0, 0.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 1.0, -1.0, 1.0] - [1.0]

Neural Network Configuration:

16 input nodes, 1 hidden layer with 9 nodes, 1 output node, RandomizedHillClimbing was the default optimizer.

Assignment 1:

I switched over my Neural Network implementation from Weka library to the ABAGAIL library.

Here is a short summary of the results from that assignment using ABAGAIL

500 training iterations	2000 training iterations	5000 training iterations
Correctly classified 120.0 instances.	Correctly classified 133.0 instances.	Correctly classified 132.0 instances.
Incorrectly classified 20.0 instances.	Incorrectly classified 7.0 instances.	Incorrectly classified 8.0 instances.
Percent correctly classified: 85.714%	Percent correctly classified: 95.000%	Percent correctly classified: 94.286%
Training time: 0.179 seconds	Training time: 0.769 seconds	Training time: 1.534 seconds
Testing time: 0.040 seconds	Testing time: 0.020 seconds	Testing time: 0.042 seconds

Part 1 - Findings:

With back propagation

# of iterations	500	1,000	2,000	5,000	10,000
RHC - accuracy	85.000%	93.571%	92.857%	95.000%	95.000%
RHC - training time	0.153s	0.298s	0.579s	1.473s	2.936s
RHC - testing time	0.015s	0.018s	0.015s	0.020s	0.022s
SA - accuracy	50.714%	82.143%	92.857%	93.571%	94.286%
SA - training time	0.148s	0.288s	0.547s	1.429s	2.955s
SA - testing time	0.013s	0.013s	0.013s	0.016s	0.016s
GA - accuracy	91.429%	93.571%	95.000%	93.571%	93.571%
GA - training time	15.100s	30.136s	60.074s	151.720s	300.609s
GA - test time	0.011s	0.010s	0.013s	0.012s	0.013s

Using forward propagation

# of iterations	400	800	2,000	5,000	10,000	20,000	200,000
RHC - accuracy	88.571%	95.000%	96.429%	96.429%	95.714%	92.857%	91.429%
RHC - training time	0.111s	0.233s	0.584s	1.385s	2.866s	5.389s	54.735s
RHC - testing time		0.015s	0.016s	0.016s	0.018s	0.019s	0.029s
SA - accuracy	50.000%	62.857%	94.286%	93.571%	95.714%	93.571%	93.571%
SA - training time	0.111s	0.222s	0.530s	1.449s	2.809s	5.446s	54.160s
SA - testing time		0.012s	0.013s	0.017s	0.014s	0.015s	0.022s
GA - accuracy	94.286%	93.571%	95.000%	92.857%	95.714%	91.429%	92.857%
GA - training time	11.075s	23.328s	55.864s	139.499s	271.043s	556.320s	5685.579s
GA - test time	0.009s	0.009s	0.009s	0.010s	0.010s	0.010s	0.011s

RHC Weights:

0.395178, -0.234528, 1.101487, -0.983102, 0.039597, 0.617959, 0.092891, 0.802494, -0.652352, -0.735453, 0.402548, 0.273986, 0.334903, -0.169873, 0.037464, 0.050946, -0.190770, -0.086778, 0.582500, -0.588829, 0.828045, -0.160971, -0.236402, 0.102323, -1.432152, 0.083076, -0.378793, -0.301411, 1.345799, 0.657633, -0.970859, -1.870704, -0.423229, -1.043819, 0.547757, 0.301632, 0.647698, -1.461752, -1.271770, -0.241187, 0.750698, 0.559141, 0.391758, 0.236943, -0.776148, -0.068342, 0.100572, -0.373041, 1.228555, -1.151828, -1.260089, 0.558581, 1.139316, 0.224655, -0.670795, -0.742411, 0.655347, -0.661998, -0.064070, -0.468882, -0.271867, 2.258243, -0.167377, 0.119522, -0.292915, -0.723943, 0.456208, -0.472881, -0.303517, -0.001593, 0.430730, -2.091236, -0.260287, 0.862998, -0.307923, -0.476425, 0.197832, 0.035503, 0.744348, -1.060493, 0.357635, 0.187779, -0.674261, -0.242501, 0.511976, -0.578881, -0.108071, -0.374289, -0.380302, -0.466126, 0.135097, 0.466162, 0.218041, 0.727011, -0.256254, 0.544615, 0.042174, 0.586091, -0.371660, 0.384520, -0.358551, 0.059968, 0.063019, 0.047538, 0.185187, 1.442374, -0.506073, -0.496257, 0.497715, -0.527962, -0.903495, -0.139178, -0.843415, 0.056667, -0.295623, 0.970518, -0.141720, -0.096625, -0.102324, -0.335422, 0.107538, 1.441608, -1.109972, -0.409639, -0.454869, -0.502785, -0.632002, -0.054046, -0.635648, 0.499270, 0.064199, -0.222635, 0.439205, -0.010209, -0.305669, 0.706101, -0.744691, -0.204613, -1.479095, 0.644289, 0.408265, 0.194531, 0.047015, -1.644729, 0.896716, 0.413566, -0.366656, -0.922042, 0.006781, -0.248885, -0.647563, 0.506341, -0.375685, -0.200943, -0.252593, 0.057048, 0.310955, -0.372170, -0.484801, -0.193820, 0.302984, 0.062893, -0.116469, -0.199235, -0.086274, -0.046248, 0.094240, 0.224027, -0.395967, 0.197060, -1.432377, 0.957166, -0.419311, -0.600161, -0.384547, -2.247588, 1.162206, -1.626895, 0.334916, -0.454936

SA Weights:

-1.164189, -0.854616, 0.271630, 1.653637, 0.241240, -0.038367, -0.298884, 0.321579, -0.959093, 0.510739, -0.105262, -0.112480, 0.451251, -0.186790, 0.286256, -0.805105, -0.184811, 0.475563, 0.316585, -0.855748, 0.678978, 0.721139, 0.533353, 0.630978, -0.709649, -0.039776, 0.237433, -1.306262, 0.257585, -0.112725, 0.905943, -1.779739, 1.057012, -0.513786, -0.177351, -0.109534, -0.607500, 1.513179, -0.155670, -1.724957, -0.304226, -0.057398, -0.173712, 0.474038, -0.392171, 1.069723, 0.524241, 0.076805, -0.251798, 0.401903, -0.142676, -1.817794, 1.877371, -1.625778, 0.787703, -0.640189, -0.430077, -0.833253, -0.315082, -0.092815, 0.596365, -0.839548, 0.488485, 0.058872, -0.300010, -1.747379, -0.090223, -1.241563, -1.121091, 0.778979, 0.023427, -1.911558, -0.363641, -1.135551, 0.710008, 0.759903, 0.953285, -1.031176, -0.559671, 1.305505, 1.098575, 0.944495, 0.905612, -0.502379, -0.669105, -0.848481, -0.663582, -0.998315, 2.317599, 0.384730, 0.432644, 0.304939, 0.192302, 0.211353, -1.581057, -1.185918, -1.363793, -1.256165, 1.817094, -0.582910, -0.340062, 1.131705, 0.157878, -0.634076, 0.194029, -0.222519, 0.706553, 0.018499, 0.201165, 2.423144, 0.284764, 1.198837, 0.459600, 0.258998, -0.739497, 1.031773, 0.454795, 0.233987, 0.333627, -1.502945, -0.334463, -0.539572, -1.065924, -0.788949, -0.720845, -1.112334, 1.584903, 1.315207, -0.054259, -2.311879, 0.196830, 0.607618, -0.258205, 0.517879, -0.762442, -0.351153, 0.220317, 0.859312, -0.650375, 1.026830, 0.630296, -1.281001, 0.225457, 0.646431,

-0.305320, -0.241796, -0.925056, 0.376916, -0.937104, 0.784127, 0.946777, 0.538347, -0.869083, -0.237183, -2.096762, 0.974272, 1.965613,
-0.704829, -1.780647, 0.211421, 0.301309, -0.487678, 0.617458, 2.075616, 1.599094, -0.525031, 0.496600, 1.004272, 0.182323, -0.824151,
1.122720, 1.004765, 0.862116, 1.294007, -0.719595, 1.122661, -0.604988, -0.493209, 1.170170, 0.186311

GA Weights:

0.331452, -0.109302, -0.571981, 0.606525, 0.272558, -0.320842, 0.275369, -0.242719, -0.522536, 0.421136, -0.643486, -0.015932, 0.292454,
-0.045380, 0.055656, -0.111445, -0.185556, 0.341178, -0.185650, -0.482141, 1.129016, -0.067580, -0.235189, 0.267395, -0.692335, -0.057760,
0.242884, -1.152568, 0.303911, -0.184146, 0.096400, -0.387090, -0.198845, -0.340219, -0.906749, 0.494778, -0.320956, 0.704649, -0.061172,
0.181533, 0.637902, -0.579207, -0.387709, -0.099783, -0.381964, -0.200918, -0.197873, 0.081749, -0.470719, 0.194959, -0.584034, 0.441544,
-0.247335, -0.844202, 1.137545, -0.155160, 0.347227, 0.255940, -0.076256, -0.070018, -0.074861, 0.018853, -0.558485, -0.195596, -0.229017,
-0.224075, -0.202376, 0.523668, 0.494344, -0.006190, 0.217558, -0.947135, -0.509837, 0.229266, -0.064263, 0.021316, -0.143636, -0.187307,
0.129140, -0.179522, 0.454523, 0.208840, -0.150948, -0.341816, -0.035352, -0.297549, -0.036736, 0.427232, -1.508281, -0.041779, 0.598977,
0.191874, -0.135934, 0.374043, -0.728594, 0.572521, -0.592314, 0.253519, -0.007724, -0.091251, 0.053228, 0.530579, -0.726641, -0.003706,
0.807431, -0.694695, -0.465804, 0.275831, -0.004494, 0.106997, 0.541083, 0.057734, 0.006167, 0.546805, 0.679984, -0.564152, -0.065460,
-0.388431, 0.026646, 0.166315, 0.250945, 0.545814, -1.112673, 0.215159, 0.507554, 0.135736, -0.350829, -0.041719, -0.725939, 0.156614,
-0.332906, -0.057236, -0.127412, 0.141716, -0.088050, 0.401788, 0.281029, -0.075113, -0.056104, -0.647896, -0.068392, 0.289596, 0.002853,
0.300217, 0.335788, 0.031900, 0.233650, 0.117896, -0.210266, -0.468129, 0.030420, -0.335319, 0.143669, -0.144452, 0.133074, -0.499675,
-0.454258, 0.038124, 0.049102, -0.036812, -0.305522, 0.239660, -0.335973, 0.332237, -0.482821, 0.205546, -0.029038, -0.496865, 0.461467,
-0.862679, 0.684890, 1.062318, 0.767398, 0.756570, -0.800006, -0.783173, -0.670823, -0.483073, -0.768998, -0.160775

Randomization algorithms analysis:

Algorithms doesn't work well with NaN doubles, using -1 for "?" value

GA execution takes a very long time to train.

GA algorithm performs well with few iterations

Simulated doesn't predict well with fewer iterations or too many.

Changing the number of iteration affect the performance of each algorithm and its execution speed.

The execution time spent when performing a test didn't change much for all the algorithms.

Part 2 - Findings:

Using random bit string in Four Peak problem:

Data Size	10	50	150	300
SA	10.0	50.0	150.0	300.0
SA time	0.000s	0.000s	0.000s	0.000s
GA	10.0	50.0	32.0	32.0
GA time	0.000s	0.000s	0.000s	0.000s
MIMIC	10.0	50.0	144.0	170.0
MIMIC time	0.000s	1.000s	10.000s	41.000s

Resources:

http://en.wikipedia.org/wiki/Mathematical_optimization

http://www.cc.gatech.edu/~isbell/tutorials/mimic-tutorial.pdf

http://www.mathworks.com/videos/using-global-search-for-optimization-problems-68768.html

Fitness functions: http://en.wikipedia.org/wiki/Fitness_function

An Introduction to Genetic Algorithms

http://www.clarkfrancis.com/docs/f_ga.html

http://www.ise.rutgers.edu/resource/research_paper/paper_05-008.pdf

Andrew W. Moore' tutorials

http://www.cs.cmu.edu/~awm/tutorials.html

http://www.autonlab.org/tutorials/hillclimb.html

http://www.autonlab.org/tutorials/hillclimb02.pdf

Tutorials on topics in machine learning

http://homepages.inf.ed.ac.uk/rbf/IAPR/researchers/MLPAGES/mltut.htm

Weka Learning algorithms: http://www.cs.waikato.ac.nz/~remco/weka_bn/node5.html

Optimal decision: http://en.wikipedia.org/wiki/Optimal_decision

http://en.wikipedia.org/wiki/Optimization_problem

http://en.wikipedia.org/wiki/Utility_maximization_problem

http://en.wikipedia.org/wiki/Hill_climbing

The Traveling Salesman with Simulated Annealing, R, and Shiny:
http://toddwschneider.com/posts/traveling-salesman-with-simulated-annealing-r-and-shiny/
Four Peaks Problem: http://www.cc.gatech.edu/classes/AY2004/cs4600_fall/projects/project3/4pks.html

Knapsack Problem: http://en.wikipedia.org/wiki/Knapsack_problem

Traveling salesman data: http://www.iwr.uni-heidelberg.de/groups/comopt/software/TSPLIB95/tsp/