Final Project: Attack Detection

Zachary Hardy (Team Captain), Patrick Behne, Peter German, Akhilesh Vijaykumar, Harpreet Singh

CSCE 633
Texas A&M University

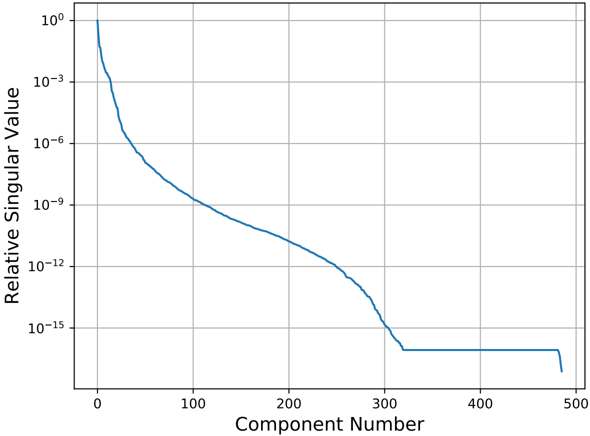
April 27, 2020

Outline

Input Transformations

- Transformations only on training data to remove bias.
- Normalization across all samples or samples sharing a time-stamp.
- Available feature-wise normalization techniques:
  - Max: Map X to $X \in \left[\frac{X_{\text{min}}}{X_{\text{max}}}, 1\right]$.
  - Standard: Map X to a zero mean and unit variance.
  - Robust: Map X by removing the median and dividing with the IQR.
- Principal Component Analysis (PCA):
  - Dimensionality reduction.
  - Fully decouple the inputs by changing to an orthogonal coordinate system.

Singular Value Decomposition of Input Data



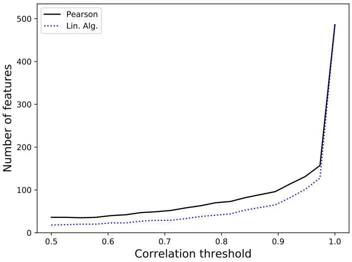
Feature Selection

- Correlation filter:
  - Pearson correlation

$$c_{i,j} = \frac{\sum_{k=1}^{N}(x_{i,k} - \bar{x}_i)(x_{j,k} - \bar{x}_j)}{\sqrt{\sum_{k=1}^{N}(x_{i,k} - \bar{x}_i)^2 \sum_{k=1}^{N}(x_{j,k} - \bar{x}_j)^2}}$$

  - Linear Algebra correlation

$$c_{i,j} = \frac{x_i \cdot x_j}{||x_i||||x_j||},$$



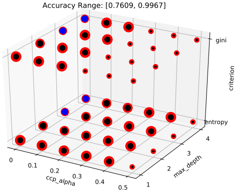
- Can filter based on:
  - feature-feature correlation
  - feature-label correlation

Previous Model Performance

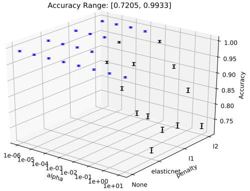
Rank	Model	Accuracy		Domain	Normalization
		Optimized	HW 3		
1	Perceptron Forest	(0.9956, 1.0000)	(0.9866, 1.0000)	Global	Any
1[a]	KNN	(0.9949, 1.0000)	(0.9814, 0.9952)	Global	Standard
1	Decision Tree	(0.9871, 1.0000)	(0.9914, 1.0000)	Global	Any
1	Perceptron	(0.9871, 0.9981)	(0.9315, 0.9600)	Global	Standard

[a] McNemar test [1].

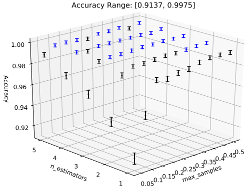
Grid Search



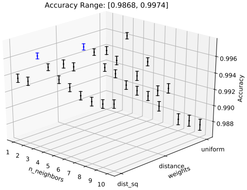
(a) Decision tree.



(b) Perceptron.



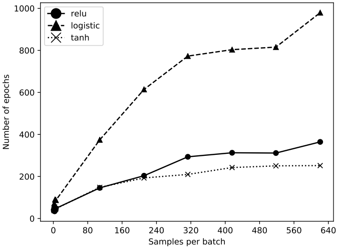
(c) Perceptron forest.

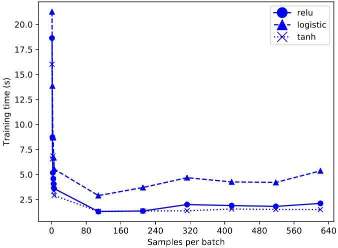


(d) KNN.

Neural Network Optimizations I

- Batch/Mini-batch/Stochastic:

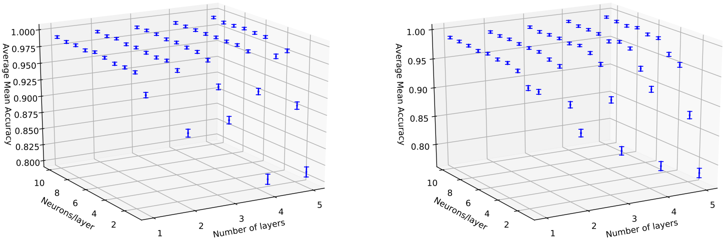




- Number of hidden neurons: 4x7
- Stochastic is the best in number of epochs till convergence
- Stochastic is the worst in terms of training time (due to large number of parameters)
- Both tanh(x) and ReLU(x) activation functions perform better than the sigmoid

Neural Network Optimizations II

- Number of layers, number of neurons per layer:



(a) Number of features: 486 (b) Number of features: 101

- Connecting multiple layers with only one neuron is not worth it.
- Above 6-7 neurons per layer, no statistical improvement observed.
- Removing redundant features has little impact on accuracy.

Best Neural Network Results

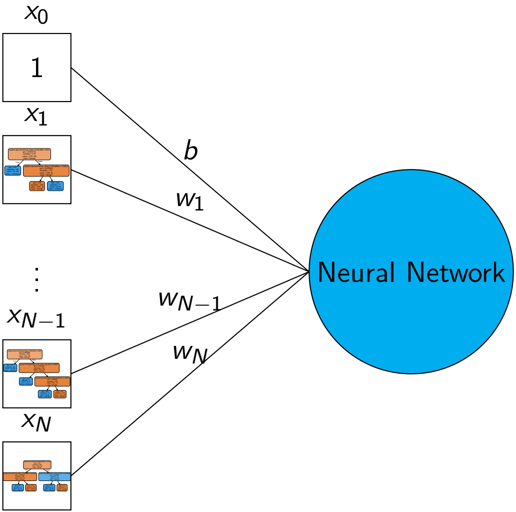
Attribute	Value
Normalization	Global Standardization
Number of layers	2
Neurons per layer	9
Mean Accuracy	0.9936
Standard Deviation	0.0026
95.00% Confidence	0.9936 ± 0.0051
Accuracy Range	(0.9885, 0.9987)
Best Accuracy	0.9968
Training Time	7.5350 s
Classification Time	0.0039 s

Slow to train and not necessarily better than other learners!

Ensemble Network Motivation

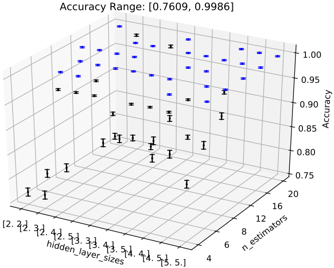
- Training neural networks is expensive.
- Employing a bagger as a filter could reduce the dimensionality of the input data and simplify the target function a neural network must learn.
- Baggers have been shown to be more accurate with unstable weak learners.
  - Unstable $\equiv$ Sensitive to training data.
  - If classifiers within an ensemble have uncorrelated outputs, the output of the ensemble will have higher accuracy than the individual classifiers.
  - The author's of [2] show this by comparing unstable decision trees to stable naive Bayes learners.
- An ensemble approach with decision trees is the appropriate design for the novel learner.

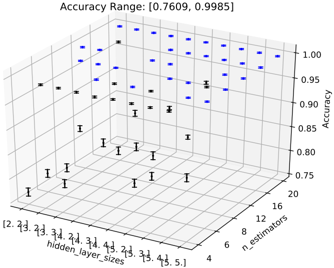
Ensemble Network Algorithm



Ensemble Network Architecture Optimization

- Number of estimators, network architecture:





- Significant reduction in overall size of neural network.
- Optimum number of hidden layers 2.
- Best Configuration:
  - Number of estimators: 8
  - Architecture: (3, 3)

Best Ensemble Network Results

Attribute	Value
Normalization	Global Standardization
Mean Accuracy	0.9979
Standard Deviation	0.0015
95.00% Confidence	0.9979 ± 0.0048
Accuracy Range	(0.9931, 1.0000)
Best Accuracy	1.0000
Training Time	2.5403 s
Classification Time	0.0242 s

Greatest mean accuracy of all models and $\approx 66\%$ reduction in training time compared to the neural network!

Model Comparisons

Rank	Model	Accuracy		Domain	Normalization
		Optimized	HW 3		
1	Perceptron Forest	(0.9956, 1.0000)	(0.9866, 1.0000)	Global	Any
1	KNN	(0.9949, 1.0000)	(0.9814, 0.9952)	Global	Standard
1	Ensemble Network	(0.9931, 1.0000)	–	Global	Any
1	Neural Network	(0.9885, 0.9987)	–	Global	Standard
1	Decision Tree	(0.9871, 1.0000)	(0.9914, 1.0000)	Global	Any
1	Perceptron	(0.9871, 0.9981)	(0.9315, 0.9600)	Global	Standard

Conclusions

- For neural networks, $\tanh(x)$ and mini-batch learning yield the minimum training times.
- The novel learner yielded comparable to better accuracy to the neural network.
- The novel learner sports a 66% reduction in training time compared to the neural network.
- Further optimizations on the novel learner architecture and exploration of PCA to come!

References

S. Raschka, "Model evaluation, model selection, and algorithm selection in machine learning."

D. W. A. D. Bazell, "Ensembles of classifiers for morphological galaxy classification," *The Astrophysical Journal*, 2001.