Background
000

Methods
0000

Results
000000000

Using Artificial Neural Networks in the *stop squark* Search

R. Brockman, J. DeVito, and R. LeVan

Department of Electrical Engineering
Rice University

Spring 2013

Background
000

Methods
0000

Results
000000000

Outline

RICE

Introduction

Overall Plan

From a large data set we try to separate a signal from a specific type of noise.

- Signal := *stop squark* event
- Noise := top quark background event
- Use SOM and Back-prop as filters

RICE
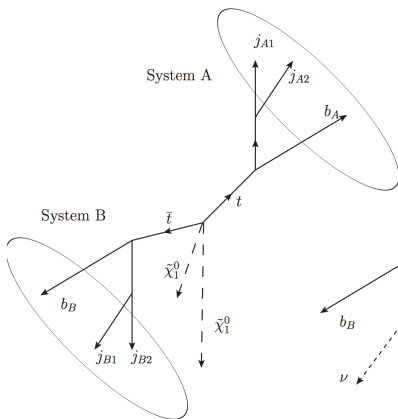
Background
○●○

Methods
○○○○

Results
○○○○○○○○○

Purpose

- Search for the *stop squark*, predicted by SUSY
- Guided by Dr. Paul Padley of Bonner Lab
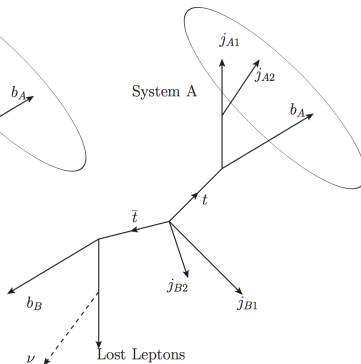- Simulated data generated by PYTHIA

Aim

Improve on the results in the paper by Dutta et al. [1], working in parallel with Onkur Sen, who is attacking the same problem using boosted decision trees.

RICE

Background
○○●

Methods
○○○○

Results
○○○○○○○○○

What is a *stop squark* Event?

Schematics of Signal

Schematics of Background

[1]

Classification Basics

From the simulated data, with the help of Onkur's code, we were able to create:

- a length 24 vector of raw data for each signal or background event

- a length 8 vector of derived variables from the raw data [1]

Previously attempted strategy:

- Thresholds on derived variables as outlined in Dutta et al. [1]

RICE

Background
000

Methods
○●○○

Results
○○○○○○○○○

Back-Propagation Filter Settings

ARCHITECTURE	
Topology	$(8 + 1_{Bias})$ - $(30 + 1_{Bias})$ - 2_{output}
Transfer Function	tanh with slope $b = 1$
LEARNING PARAMETERS	
Initial weights	$w \sim U[-0.1, 0.1]$
Learning rate, $\gamma(t)$	$\gamma(t) = 0.01(1 - 0.0001)^t$
Momentum, α	$\alpha = 0.3$
Epoch size	$K = 1$
Stopping criteria	learning step > 100,000
Error measure (Err)	RMSE
Monitoring frequency (m)	1,000 Learning Steps
INPUT/OUTPUT SCALING	
Input Scaling	(-0.9,0.9)
Output Scaling	(-0.9,0.9)
PERFORMANCE EVALUATION	
Accuracy measure (Acc_X)	Significance = $\frac{S}{\sqrt{B}}$

RICE

Background
000

Methods
○○●○

Results
○○○○○○○○○

Self-Organizing Map Filter

ARCHITECTURE	
Topology	10 x 10
LEARNING PARAMETERS	
Initial weights	$w \sim U[-0.1, 0.1]$
Learning rate, $\gamma(t)$	$\gamma(t) = 0.3(1 - 0.00001)^t$
Neighborhood, $\sigma(t)$	$\sigma(t) = 1.5 + 3.5(1 - 0.00001)^t$
Epoch size	$K = 1$
Stopping criteria	learning step > 100,000
Monitoring frequency (m)	1,000 Learning Steps
INPUT/OUTPUT SCALING	
Input Scaling	Angles in Degrees, Otherwise None
Output Scaling	None
PERFORMANCE EVALUATION	
Accuracy measure (Acc_X)	Significance $= \frac{S}{\sqrt{B}}$

RICE

Significance: A Filter Comparison Metric

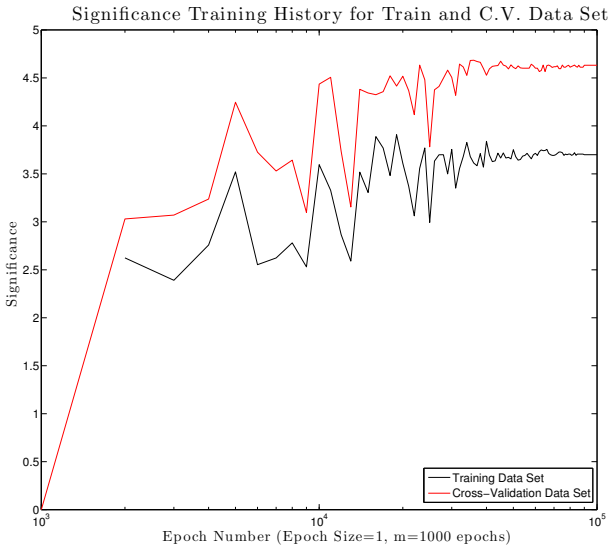- A measure of discovery confidence based on a certain number of high energy particle collisions

$$\text{significance} = \frac{N_{\text{Signal}}}{\sqrt{N_{\text{Background}}}}$$

$N_{\text{Signal}} :=$ number of signal events that come through the filter
$N_{\text{Background}} :=$ number of noise events that come through the filter

- Used commonly in physics
- Every significance measure in the Results Section is based on a standard number of collisions

RICE

Background
○○○

Methods
○○○○

Results
●○○○○○○○○○

Back-Propagation Results for Derived Variables

Significance Training History for Train and C.V. Data Set

Background
ooo

Methods
oooo

Results
o●ooooooo

SOM for Derived Variables

SOM Signal and Noise Density plots with weights

Signal

Background
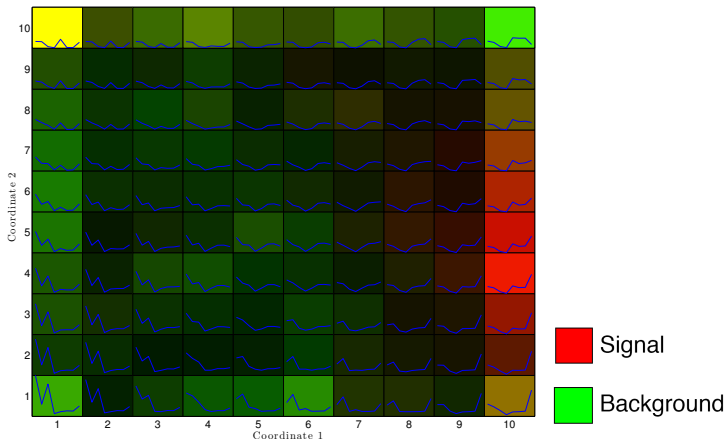
SOM for Derived Variables

SOM Normalized Signal to Noise Ratios

Background
○○○

Methods
○○○○

Results
○○○●○○○○○

SOM followed by BP for Derived Variables

Significance Training History for Train and C.V. Data Set

Background
○○○

Methods
○○○○

Results
○○○○●○○○○

SOM on Raw Data

SOM Signal and Noise Density plots with weights

Signal

Background

Background
000

Methods
0000

Results
00000●000

SOM on Raw Data

SOM Normalized Signal to Noise Ratios

Background
○○○

Methods
○○○○

Results
○○○○○○●○○

Significance Analysis

METHOD	TEST SET SIGNIFICANCE
Thresholding [1]	1.93σ
No Filter	2.62σ
Back-Propagation	3.79σ
Self-Organizing Map for Derived Variables	3.69σ
Self-Organizing Map and Back-Propagation	4.36σ
Self-Organizing Map for Raw Data	2.62σ
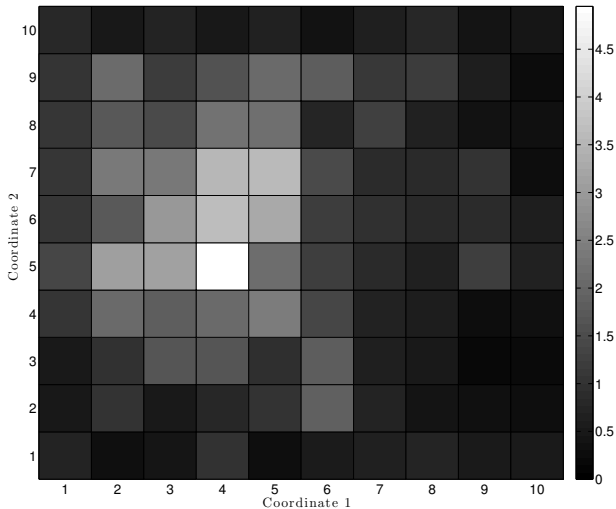
Analysis

- Back-Propagation and the SOM using the 8 derived variables produced interesting results

- The SOM using 24 raw variables did not

- The two-stage SOM and BP process was slightly superior to either method alone

- We suspect the success of the 8 derived variables is due to the jet angle invariance of these parameters

RICE

Background
○○○

Methods
○○○○

Results
○○○○○○○○●

Next Steps

- Determine a method of aligning the 24 non-derived variables
- Further Experimentation with Training Parameters
- Run a second SOM on the SOM cells with high signal to noise ratios

RICE

References

[1] Dutta, B.; Kamon, T.; Kolev, N.; Sinha, K.; Wang, K. *High Energy Physics - Phenomenology* **2012**,