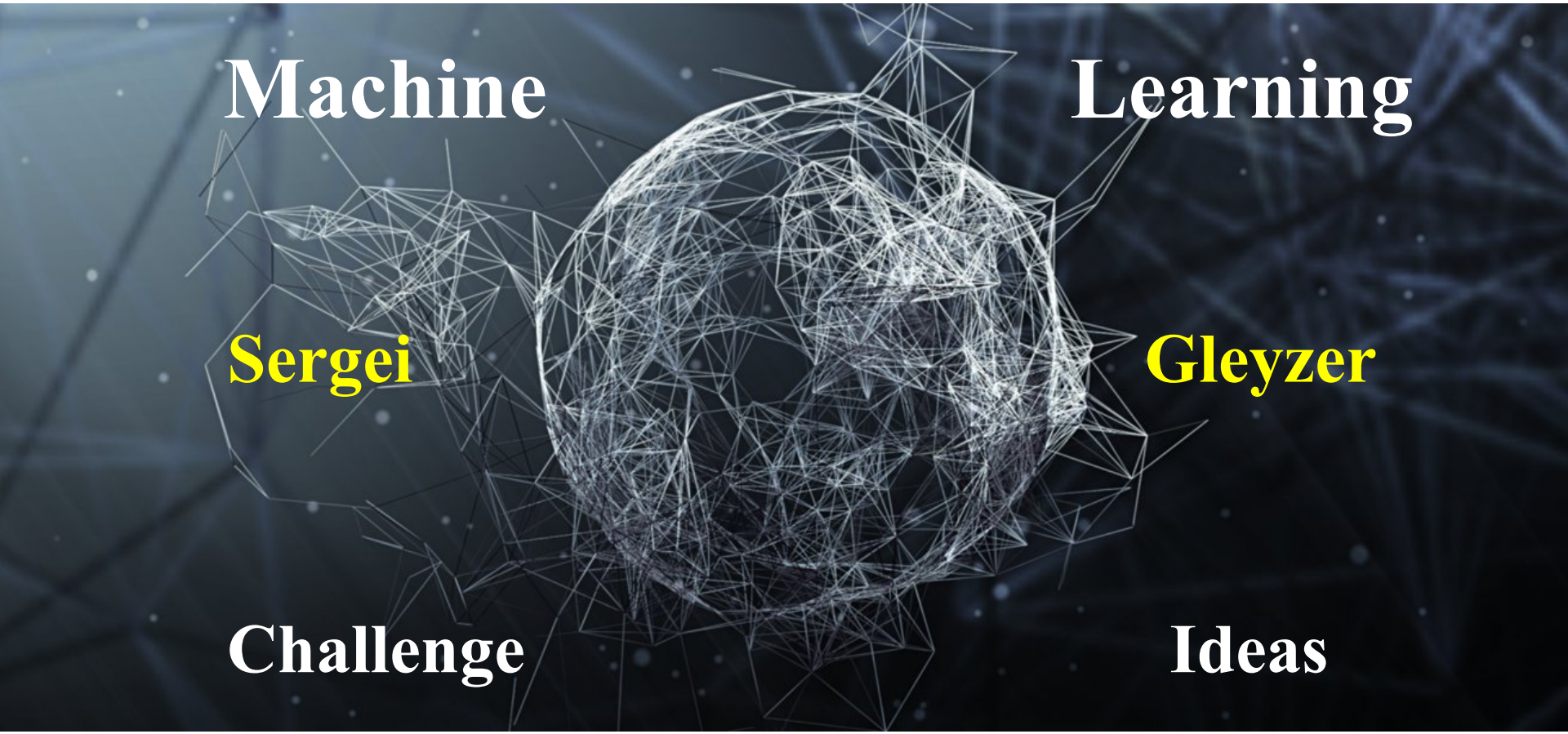


Inter-Experimental Machine Learning Working Group Meeting

Dec. 4, 2015

Machine Learning Challenge Ideas

Outline

- Classification/Regression in HEP
- Regression Challenge
- Multi-target regression

Motivation

Common machine learning problems in HEP:

- **Classification**
  - New physics event or background?
    - Past challenges are of this kind
- **Regression/Clustering**
  - How to best model particle energy based on observables/detector measurements
    - Lots to gain from new ideas in this area (our own attempts show improvement)

Classification

Distinguish **f(x)**, **g(x)** using Training set of observations

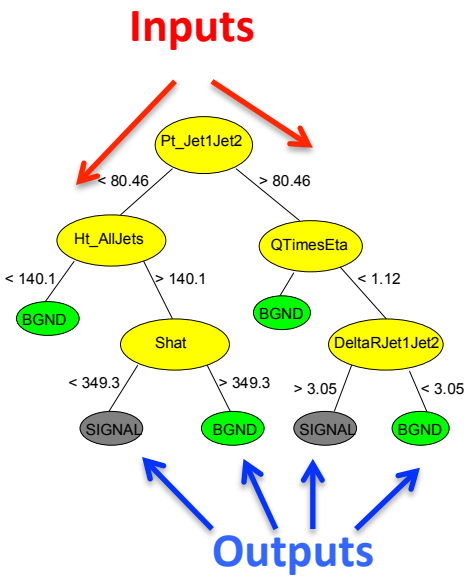
{**inputs** , **outputs**}

Pass **observations** to a learning algorithm

neural network, decision tree

that produces **outputs** in response to **inputs**

Use another set of observations to evaluate



Function Estimation

Regression

Inputs: Training examples $\{<x^{(i)}, y^{(i)}>\}$ of unknown function f. $x^{(i)}, y^{(i)}$:

Output: hypothesis h that best approximates target function f

Regression

Regression

- modify the evaluation criteria used in the induction algorithm
  - from maximum **separation** gain
  - to minimal **variance**
    - **Simple change, but there are many subtleties**

Regression Trees

Regression/Clustering:

Obtain clusters with decision tree induction:

- **minimal intra-cluster distance**
  - between examples from the same cluster
- **maximal inter-cluster distance**
  - between examples from different clusters
- In classification trees distance metric is **class enthropy**
- In regression trees it is **variance**

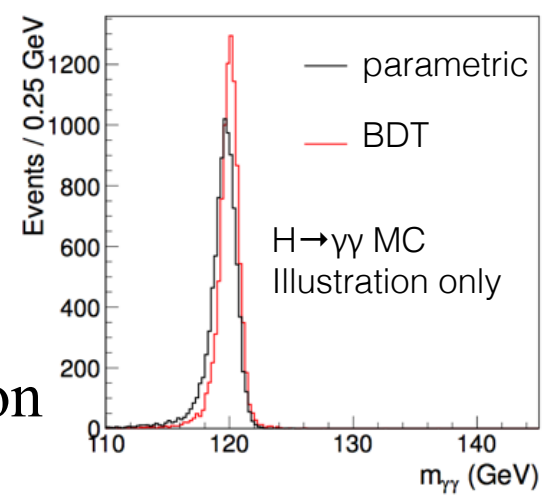
Photon Energy

Improve photon energy measurement by applying regression

Inputs: photon (electromagnetic) shower information, photon coordinates, median event energy

Target Output: E_{REC}/E_{TRUE}

10-30% improvement in resolution depending on energy and detector region



Regression Challenge

Setup a challenge focused on **regression**

Improvements in particle/detector measurements are worthwhile

- Bottom line: translates to significance/reduction in error measurement improvements

Already used in photon/electron/b-jet energy measurements. Can we do better?

Multi-Target Regression

Multiple Targets

How to estimate a number of functions or targets at once

- Preserving the internal **relationships** between input features

Use cases:

- fast detector simulations, multivariate transformations for individual particles, many others not yet explored

Example

14 input variables {a, b, c, d…}

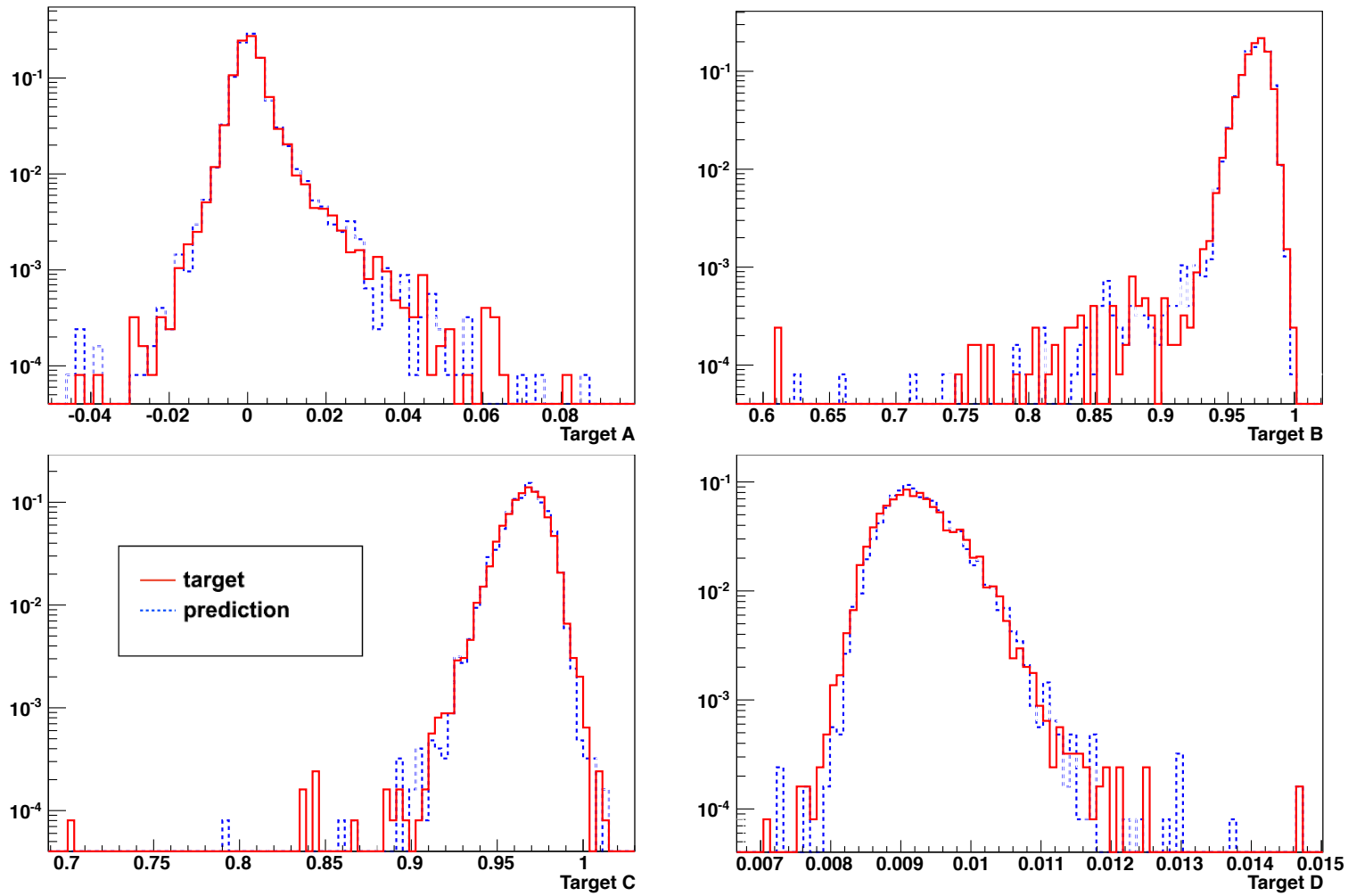
– 4 of them strongly correlated

14 target outputs to estimate {A, B, C, D…}

– 4 of them strongly correlated

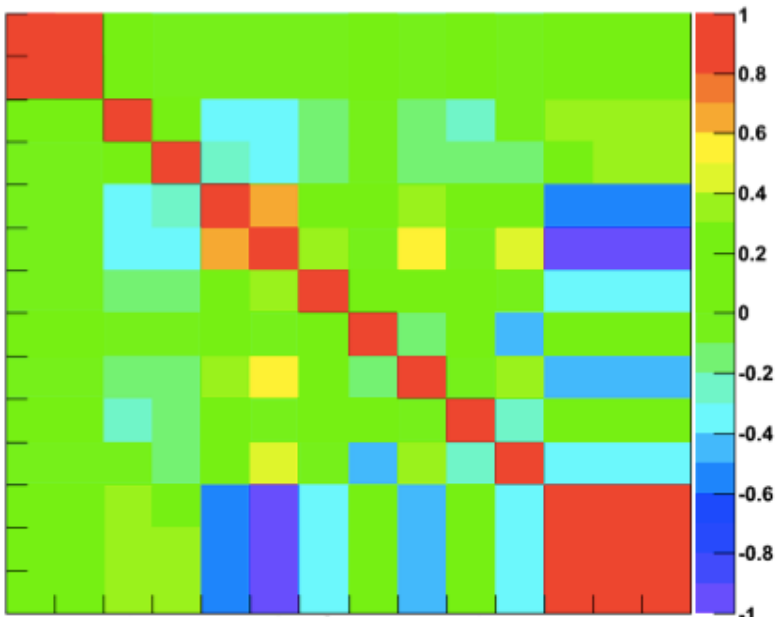
Challenge: build a predictive model to describe simultaneously all the outputs {A,B,C,D…}, provided a corresponding set of inputs.

Illustrative Example #2

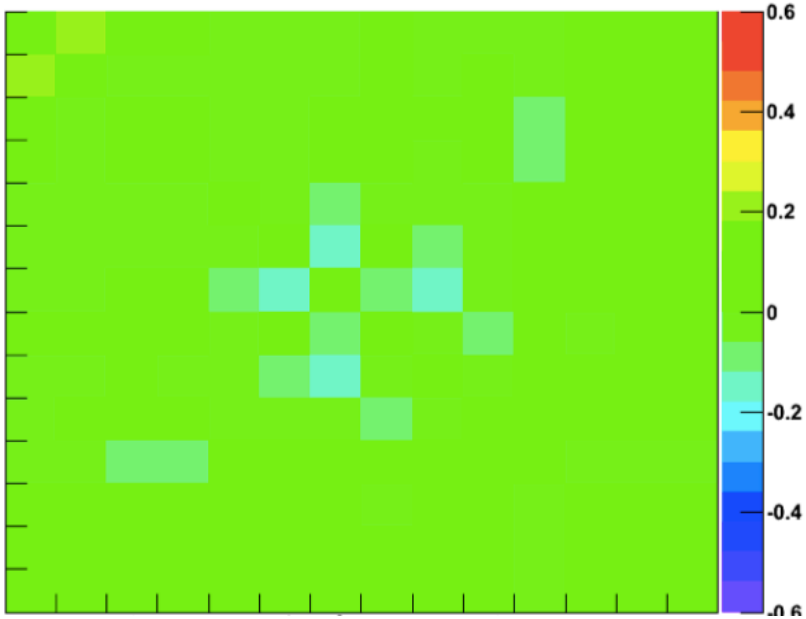


Correlations

Target Correlations



Prediction-Target Difference



Very close to Zero

Summary

- **Past** HEP ML challenges focused on classification
  - avoid repetition
- **Regression** challenge
  - Potential further gains in particle energy measurements
- **Multi-target** regression also useful and interesting to study