Multivariate Jet Calibration Using Neural Networks

-Dayton Grogan

-Duke University
Large-Radius Jets

Why do we care?
• Provides simplified event reconstruction
  • Jet energy → mass scale of the process
  • Jet mass → identity of particle
• Understanding scale and resolution is important!
Energy Calibration

- Calibration from MC, \( R \approx \frac{\text{True energy}}{\text{Reconstructed energy}} \)

- \( \text{true } pT \approx R(Jet \ pT, \eta) \times \text{reconstructed } pT \)

- Does more information give more precision?

- Limitations
  - not very practical

- Use Neural Networks!
Neural Networks

Why Neural Nets?
• Accommodates many variables
• Easily updatable

Network Training
• Input: simulated data
• Output: calibration factor R
• loss function: $L2 = (\frac{true\ p_T}{reco\ p_T} - R)^2$

Simulated data sample:
• simulated di-jet events
• $p_T > 200$ GeV, $|\eta| < 2.0$
• matched geometrically to true jet
Network Configuration

Many options!
- Number of layers
- Number of nodes (neurons)
- Activation functions
- Propagation algorithms

Does it matter?
- Yes!
- Effects on runtime and convergence

Epoch: one forward pass and one backward pass of all training examples
Impact of Structure: Layers

- Multiple layers may improve convergence
- Runtime increased by factor of 2.5/epoch
- Still effective to use a single layer
Impact of Structure: Nodes

- Performance not correlated with nodal number
- More input variables requires more nodes

ATLAS work in progress
Impact of Structure: Activation Function

- Relu (rectified linear unit)
- Sigmoid and tanh are more comparable

**ATLAS work in progress**

- Network Optimization
  - Large space of possible configurations for the network and training parameters:
    - **Network Parameters**
      - Number of layers: 1-3
      - Neurons per layer: 10-50
      - Activation function: relu, tanh or sigmoid
    - **Optimization Parameters**
      - Loss function: $L_2 = \| p^T, reco - p^T, truth \|$  
  - Propagation Algorithm:
    - Gradient Descent
    - Momentum
    - Adam
  - Training rate: Large variance

**Graph**

- Log(mean square error) vs Epoch
- Sigmoid
- tanh
- Relu

**Equations**

- **Sigmoid**
  \[
  f(x) = \frac{1}{1 + e^{-x}}
  \]

- **Tanh**
  \[
  \tanh(x) = \frac{2}{1 + e^{-2x}} - 1
  \]

- **RelU**
  \[
  f(x) = \begin{cases} 
  0 & \text{for } x < 0 \\
  x & \text{for } x \geq 0 
  \end{cases}
  \]
Final Network Structure

Network Parameters

• Number of layers: 1

• Number of nodes per layer: 20-50

• Activation function: Tanh

• Propagation algorithm: Adam
Preliminary Results

So, how are we doing?

- Successfully calibrated!
- Convergence occurs within 10,000 epochs and takes roughly 2 hours.
Preliminary Results

How do we compare?

• NN looks very similar to by-hand JES calibration
Future Directions

- Formalize optimization strategy and find optimal $p_T$ calibration
- Repeat studies with new (jet substructure) input variables and less generic jets
- Expand strategy to calibrate other observables
  - particularly jet mass!
Backup Slides
Back up Slides: Layers

<table>
<thead>
<tr>
<th>Layers</th>
<th>Final Training MSE</th>
<th>Final Validation MSE</th>
<th>Training time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>378.272</td>
<td>67.8496</td>
<td>107 sec</td>
</tr>
<tr>
<td>2</td>
<td>362.789</td>
<td>83.2652</td>
<td>189 sec</td>
</tr>
<tr>
<td>3</td>
<td>360.262</td>
<td>85.7885</td>
<td>250 sec</td>
</tr>
<tr>
<td>4</td>
<td>360.221</td>
<td>85.7614</td>
<td>290 sec</td>
</tr>
</tbody>
</table>

One data file (JZW7) and 20 nodes

ATLAS work in progress
Backup Slides: Layers

- Convergence is comparable among the different layers.
- Only see difference when lower numbers of nodes are used.
Backup Slides: Propagation algorithm

- although there are strong early fluctuations away from best calibration, Adam converges fastest

- Other algorithms are smoother, but may be susceptible to getting stuck in a local minimum.

Adadelta, Adagrad, and Adam

ATLAS work in progress
Backup Slides: Calibration Comparison

Comparison of Final Response Functions (Normalized)

Legend:
- No Calibration
- New Calibration
- Old Calibration
- Target at 1

ATLAS work in progress