**Introduction**

- RBM training with CD works well when data is sparse (black background)
- Training is much worse with inverted images
- We present a simple and effective solution to this problem
- New algorithm requires the addition of 3 lines of code
- Data normalization improves RBM training and should *always* be used

**Algorithm**

**Algorithm 1** Contrastive Divergence Training of RBM on zero-meaned data.

1. Subtract data mean from all training data vectors: \( z = v - \mu \)
2. Compute hidden activations and sample using Eq. 6.
3. Calculate the MLE gradient of the positive phase.
4. Compute reconstructions \( \hat{z} = \hat{v} - \mu \)
5. Compute reconstructions \( \hat{z} = \hat{v} - \mu \)
7. Calculate the MLE gradient of the negative phase.
8. Approx. gradient = positive phase gradient − negative phase gradient.
9. Update the parameters.
10. Modify hidden biases: \( c_{new} \leftarrow c - \mu^TW \)

**AIS estimation of log-prob in nats: test (training)**

<table>
<thead>
<tr>
<th>Dataset</th>
<th>FPCD-30</th>
<th>ZM-30</th>
<th>FPCD-1000</th>
<th>ZM-1000</th>
</tr>
</thead>
<tbody>
<tr>
<td>MNIST</td>
<td>-96 (-96)</td>
<td>-94 (-94)</td>
<td>-84 (-81)</td>
<td>-84 (-81)</td>
</tr>
<tr>
<td>Neg. MNIST</td>
<td>-110 (-110)</td>
<td>-96 (-96)</td>
<td>-87 (-85)</td>
<td>-84 (-81)</td>
</tr>
</tbody>
</table>

**Conclusions**

- Extremely simple way to improve training
- Zero-mean works well with CD, PCD, and FPCD
- Leads to models with higher log-probs
- Learns sparser features, which are better for classification

Tech Report: www.cs.toronto.edu/~tang