Learning Feature Hierarchies by Learning Deep Generative Models

Ruslan Salakhutdinov
BCS and CSAIL, MIT
Talk Outline

1. Discriminative vs. Generative.
2. Deep Belief Networks (DBN’s).
3. Kernel Learning with DBN’s.
Many real-world applications: high-dimensional, highly-structured data

Large supply of unlabeled data and a very limited amount of labeled data.

Applications such as information retrieval and machine vision are examples where unlabeled data is readily available.
Discriminative vs. Generative

- Given a set of $i.i.d$ training samples $\{x_l, y_l\}$.

- Discriminative models model $p(y_l|x_l; \theta)$ directly (logistic regression, Gaussian process, SVM’s).

- A large supply of unlabeled data $\{x_u\}$.

- Need to make some assumptions about the input data $\{x_u\}$.

- Otherwise unlabeled data is of no use.
Discriminative vs. Generative

Key points of learning deep generative models:

- Learn probabilistic model $p(x_u; \theta)$.

- Use learned parameters $\hat{\theta}$ to initialize a discriminative model $p(y_n | x_n; \hat{\theta})$ (neural network).

- Slightly adjust discriminative model for a specific task.

No knowledge of subsequent discriminative task during unsupervised learning. Most of the information in parameters comes from learning a generative model.
Building Block: RBM’s

Restricted Boltzmann Machines: 2-layer modules.

Visible stochastic binary units \( v \) are connected to hidden stochastic binary feature detectors \( h \):

\[
P(v, h) = \frac{1}{\mathcal{Z}} \exp \left[ \sum_{ij} v_i h_j W_{ij} \right].
\]

Markov Random Fields, Log-linear Models, Boltzmann machines.
Unsupervised Learning of DBN’s

Deep Belief Networks.

Greedy, layer-by-layer learning:

• Learn and Freeze $W^1$.
• Sample $h^1 \sim P(h^1|v; W^1)$.
  Treat $h^1$ as if it were data.
• Learn and Freeze $W^2$.
• ...

Learn high-level representations.
Kernel Learning

Deep models can be used to learn kernel function for many discriminative methods: SVM’s, kernel regression, Gaussian processes.

- Learn a deep generative model of $p(x; W)$ in an entirely unsupervised way: DBN or Deep Boltzmann machines.
- Use this deep generative model to initialize a kernel function $K(x, y; W)$, parametrized by $W$.
- Use backpropagation to discriminatively fine-tune parameters $W$ of the kernel.
Learning Covariance Kernel

- Initialize covariance function of the Gaussian Process parameterized by \( \theta = \{ \alpha, \beta, W \} \):

\[
K_{nm} = \alpha \exp \left( - \frac{1}{2\beta} \left\| F(x^n; W) - F(x^m; W) \right\|^2 \right).
\]

- Learn \( \theta \) by maximizing the marginal likelihood.
Regression Task

Predicting the orientation of a face patch.

-66.84  43.48  -57.14  14.22  -35.75  30.01

• Labeled Training Data:
  Input: 1000 labeled training patches  Output: orientation

• Labeled Test Data:
  Input: 1000 labeled test patches  Predict: orientation of new people.

• Gaussian Processes with exponential kernel achieves a RMSE of 16.36° (±0.45°).
Regression Task

-66.84  43.48  -57.14  14.22  -35.75  30.01  Unlabeled

- Additional Unlabeled Training Data: 12000 face patches.
- Learn a DBN: 784-1000-1000.
- Features were extracted without knowledge of the final task.

The same GP on the top-level features: RMSE 11.22°.
Learn the covariance function of GP: RMSE 6.42°.
Nonlinear Neighbourhood Component Analysis.

**Unsupervised learning:** Learn a non-linear transformation of the input space.

**Discriminative learning:** Using labels, optimize to make KNN perform well in the low-dimensional feature space.
The 2-dimensional mappings for MNIST and 20-newsgroup datasets.
Deep Boltzmann Machines

\[ P(v) = \sum_{h^1, h^2, h^3} \frac{1}{Z} \exp \left[ v^\top W^1 h^1 + h^1^\top W^2 h^2 + h^2^\top W^3 h^3 \right]. \]

Deep Generative Model: Markov random field with hidden units.

Fast greedy initialization.

Bottom-up + Top-down.

Unsupervised learning of high-level representations.

Labeled data is used to only slightly fine-tune the model.
Discriminative fine-tuning: test error of 7.2%.
SVM’s get 11.6%, logistic regression gets 22.5%.
Image Completion
Thank you.