ECE 599/692 – Deep Learning

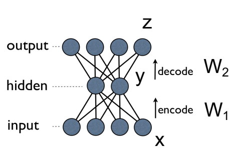
Lecture 9 – Autoencoder (AE)

Hairong Qi, Gonzalez Family Professor
Electrical Engineering and Computer Science
University of Tennessee, Knoxville
http://www.eecs.utk.edu/faculty/qi
Email: hqi@utk.edu

Outline

- Lecture 9: Points crossed
  - General structure of AE
  - Unsupervised
  - Generative model?
  - The representative power
    - Basic structure of a linear autoencoder
    - Denoising autoencoder (DAE)
  - AE in solving overfitting problem
- Lecture 10: Variational autoencoder
- Lecture 11: Case study

General structure



$$\mathbf{y} = f_{\theta_1}(\mathbf{W}_1\mathbf{x} + \mathbf{b}_1)$$

$$\mathbf{z} = g_{\theta_2}(\mathbf{W}_2\mathbf{y} + \mathbf{b}_2)$$

$$\theta_1 = \{\mathbf{W}_1, \mathbf{b}_1\}, \theta_2 = \{\mathbf{W}_2, \mathbf{b}_2\}$$

$$\theta_1^*, \theta_2^* = \arg\min_{\theta_1,\theta_2} \frac{1}{n}\sum_{i=1}^{n} L(\mathbf{x}^{(i)}, \mathbf{z}^{(i)}) = \arg\min_{\theta_1,\theta_2} \frac{1}{n}\sum_{i=1}^{n} L(\mathbf{x}^{(i)}, g_{\theta_2}(f_{\theta_1}(\mathbf{x}^{(i)})))$$

$$L(\mathbf{x}, \mathbf{z}) = \frac{1}{2}\|\mathbf{x} - \mathbf{z}\|_2^2$$

$$L_H(\mathbf{x}, \mathbf{z}) = -\sum_{i=1}^{n}[\mathbf{x}_i \log \mathbf{z}_i + (1 - \mathbf{x}_k)\log(1 - \mathbf{z}_k)]$$

$$\theta_1^*, \theta_2^* = \arg\min_{\theta_1,\theta_2} E_{q(X)}[L_H(X, g_{\theta_2}(f_{\theta_1}(X)))]$$

Generative Model

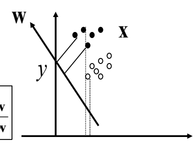
- The goal is to learn a model *P* which we can sample from, such that *P* is as similar as possible to P_{gt}, where P_{gt} is some unknown distribution that had generated examples *X*
- The ingredients
  - Explicit estimate of the density
  - Ability to sample directly

4

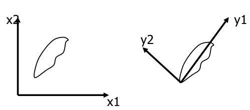
Discrimination vs. Representation of Data

- Best discriminating the data
  - Fisher's linear discriminant (FLD)
  - NN
  - CNN

$$J(\mathbf{w}) = \frac{|\tilde{m}_1 - \tilde{m}_2|^2}{\tilde{s}_1^2 + \tilde{s}_2^2} = \frac{|\mathbf{w}^T(\mathbf{m}_1 - \mathbf{m}_2)|^2}{\mathbf{w}^T\mathbf{S}_1\mathbf{w} + \mathbf{w}^T\mathbf{S}_2\mathbf{w}} = \frac{\mathbf{w}^T\mathbf{S}_B\mathbf{w}}{\mathbf{w}^T\mathbf{S}_W\mathbf{w}}$$



- Best representing the data
  - Principal component analysis (PCA)



$$\mathbf{x} = \sum_{i=1}^{m} y_i \mathbf{b}_i + \sum_{i=m+1}^{d} y_i \mathbf{b}_i \approx \sum_{i=1}^{m} y_i \mathbf{b}_i + \sum_{i=m+1}^{d} \alpha_i \mathbf{b}_i$$

Error: $\Delta\mathbf{x} = \sum_{i=m+1}^{d} (y_i - \alpha_i)\mathbf{b}_i$

5

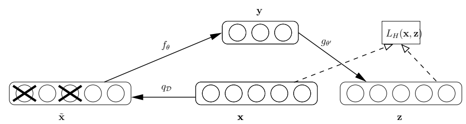
PCA as Linear Autoencoder

Raw data (X_{nxd}) → covariance matrix (Σ_X) → eigenvalue decomposition (λ_{dx1} and E_{dxd}) → principal component (P_{dxm}) → $Y_{nxm} = X_{nxd} * P_{dxm}$

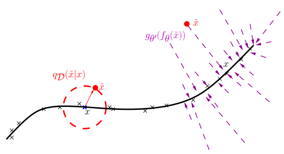
6

Denoising Autoencoder (DAE)

AICIP RESEARCH



$$\theta_1^*, \theta_2^* = \arg\min_{\theta_1,\theta_2} E_{q(X,\tilde{X})}[L_H(X, g_{\theta_2}(f_{\theta_1}(\tilde{X})))]$$



[DAE:2008]

THE UNIVERSITY OF TENNESSEE

7

The Two Papers in 2006

AICIP RESEARCH

- [Hinton:2006a] G.E. Hinton, S. Osindero, Y.W. Teh, “A fast learning algorithm for deep belief nets,” Neural Computation, 18(7):1527-1554, 2006.
- [Hinton:2006b] G.E. Hinton, R.R. Salakhutdinov, “Reducing the dimensionality of data with neural networks,” Science, 313:504-507, July 2006.

THE UNIVERSITY OF TENNESSEE

8

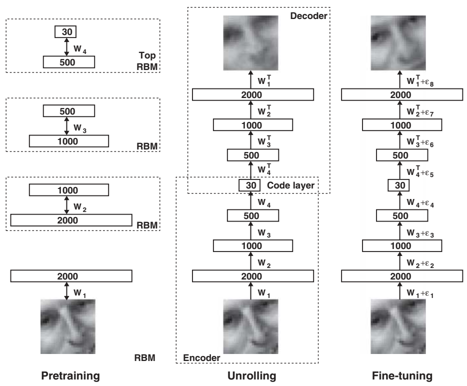
Techniques to Avoid Overfitting

AICIP RESEARCH

- Regularization
  - Weight decay or L1/L2 normalization
  - Use dropout
  - Data augmentation
- Use unlabeled data to train a different network and then use the weight to initialize our network
  - Deep belief networks (based on restricted Boltzmann Machine or RBM)
  - Deep autoencoders (based on autoencoder)

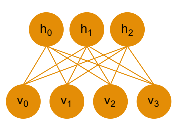
THE UNIVERSITY OF TENNESSEE

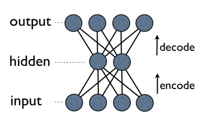
9



[Hinton:2006b]

RBM vs. AE

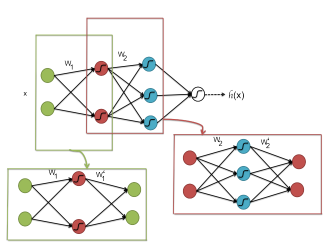




Stochastic vs. Deterministic
p(h,v) vs. Gradient of log-likelihood

AE as Pretraining Methods

- Pretraining step
  - Train a sequence of shallow autoencoders, greedily one layer at a time, using unsupervised data
- Fine-tuning step 1
  - Train the last layer using supervised data
- Fine-tuning step 2
  - Use backpropagation to fine-tune the entire network using supervised data



AICIP RESEARCH

Recap

- General structure of AE
- Unsupervised
- Generative model?
- The representative power
  - Basic structure of a linear autoencoder
  - Denoising autoencoder (DAE)
- AE in solving overfitting problem

THE UNIVERSITY OF TENNESSEE

13