NPTEL ONLINE CERTIFICATION COURSES

Course Name: Deep Learning

Faculty Name: Prof. P. K. Biswas
Department : E & ECE, IIT Kharagpur

Topic

Lecture 32: Autoencoder Variants

CONCEPTS COVERED

Concepts Covered:

- Autoencoder
  - Undercomplete Autoencoder
  - Autoencoder vs. PCA
  - Deep Autoencoder Training
  - Sparse Autoencoder
  - Denoising Autoencoder
  - Contractive Autoencoder
  - Convolution Autoencoder

Sparse Autoencoder

Sparse Autoencoder

- Interesting features can be learnt even when number of nodes in the hidden layer is large.
- Introduce sparsity constraint on the hidden layer nodes that penalize activations within a layer.
- Network learns encoding-decoding that relies on activating a small number of neurons.

Regularizing Activations not the Weights

Sparsity Constraint

$a_j^h \rightarrow$ Activation of j^{th} Neuron in hidden layer h

$a_j^h \rightarrow 1 \Rightarrow$ Neuron is active

Average activation $\rightarrow \hat{\rho}_j = \frac{1}{m}\sum_{i=1}^{m} a_j^h(x_i)$

Constraint $\rightarrow \hat{\rho}_j = \rho$

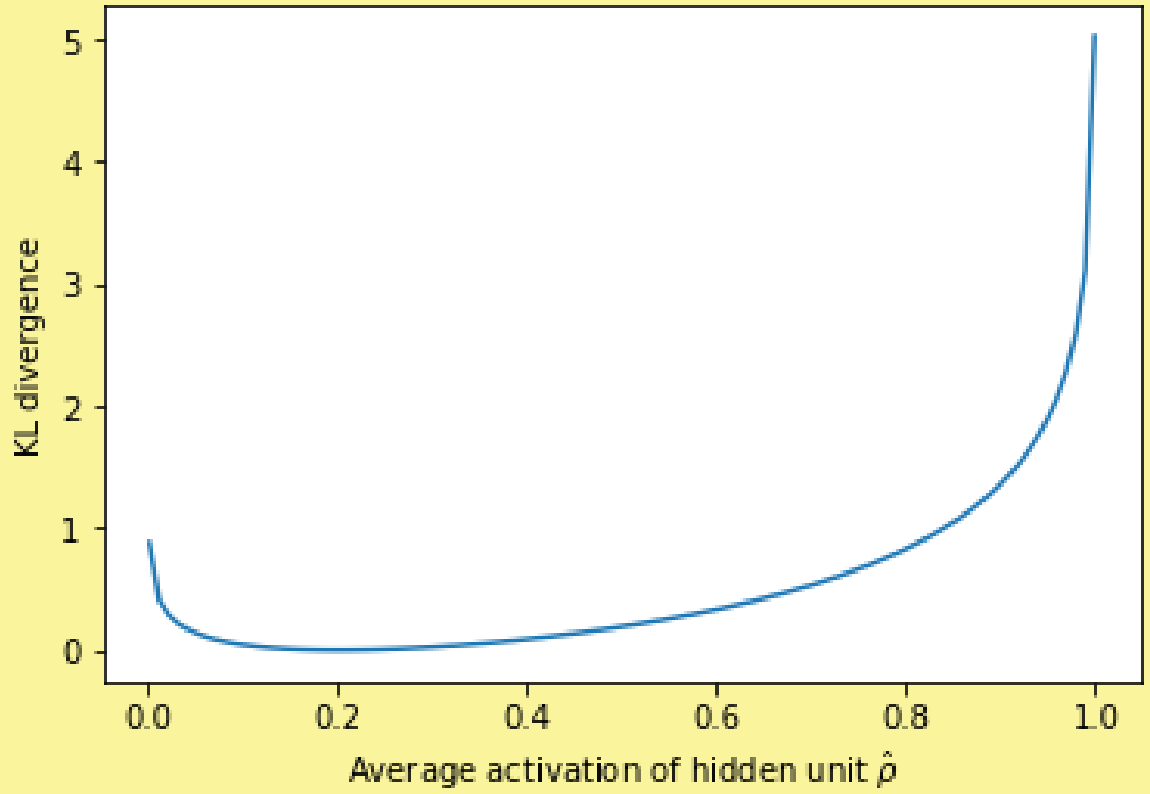
$\rho \rightarrow$ sparsity parameter (typically a small value)

Sparsity Constraint

$$\text{Regularizer}: \sum_{j=1}^{N_h}\left[\rho \log\frac{\rho}{\hat{\rho}_j} + (1-\rho)\log\frac{1-\rho}{1-\hat{\rho}_j}\right] \Rightarrow \sum_{j=1}^{N_h} KL(\rho \,\|\, \hat{\rho}_j)$$

$$J_{sparse}(W) = L(X, \hat{X}) + \lambda \sum_{j} KL(\rho \,\|\, \hat{\rho}_j)$$

KL Divergence



https://www.jeremyjordan.me/autoencoders/

Sparsity Constraint

$$\delta_i^k = O_i^k (1 - O_i^k) \sum_{j=1}^{M_{k+1}} \partial_j^{k+1} W_{ij}^{k+1}$$

$$\delta_i^k = O_i^k (1 - O_i^k) \left[\left(\sum_{j=1}^{M_{k+1}} \partial_j^{k+1} W_{ij}^{k+1} \right) + \lambda \left(-\frac{\rho}{\hat{\rho}_i} + \frac{1-\rho}{1-\hat{\rho}_i} \right) \right]$$

Denoising Autoencoder

Denoising Autoencoder

- The Autoencoder learns a generalizable encoding-decoding scheme.
- An approach:- while training use corrupt data as input but output as uncorrupted original data.
- The model can not memorize the training data as input and target output is not same any more
- The Model learns a vector field to map the input data towards a low dimensional manifold.

NPTEL ONLINE CERTIFICATION COURSES

Thank you