

CS-667 Advanced Machine Learning

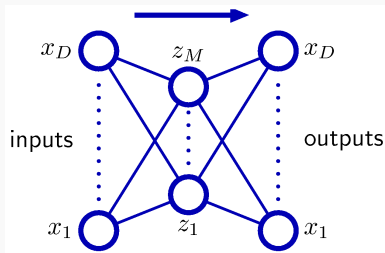
Nazar Khan

PUCIT

Lectures 25
Autoassociative Neural Networks
May 25, 2016

Autoassociative Neural Networks

- ▶ Neural nets learn the mapping from inputs \mathbf{x}_n to targets \mathbf{t}_n .
- ▶ If target is set to the input vector itself ($\mathbf{t}_n = \mathbf{x}_n$), the network learns to associate each input vector with itself.
- ▶ This is called an *autoassociative mapping* and the network is called an *autoassociative network*.



- ▶ Autoassociative nets perform unsupervised learning.
- ▶ For $M < D$, hidden layer output $\mathbf{z} \in \mathbb{R}^M$ represents *dimensionality reduction*.

Two Layer Autoassociative Nets

Equivalence with PCA

- ▶ It can be proven that for two layer autoassociative nets, outputs of the M hidden neurons correspond to projection of x onto the M -dimensional subspace spanned by the first M principal components of the data.
- ▶ This is true when activation functions of the hidden neurons are linear as well as when they are non-linear.
- ▶ Weights of hidden neurons form the basis set that spans the principal subspace.
- ▶ However, they need not be orthogonal or normalised.
- ▶ There is *no advantage* over standard PCA methods that guarantee
 - ▶ correct solution
 - ▶ in finite time
 - ▶ ordered eigenvalues
 - ▶ orthonormal eigenvectors.

Multilayer Autoassociative Nets

Nonlinear PCA

