Neural networks

Autoencoder - definition

UNSUPERVISED LEARNING

Topics: unsupervised learning

- Unsupervised learning: only use the inputs $\mathbf{x}^{(t)}$ for learning

  ‣ automatically extract meaningful features for your data

  ‣ leverage the availability of unlabeled data

  ‣ add a data-dependent regularizer to trainings

- We will see 3 neural networks for unsupervised learning

  ‣ restricted Boltzmann machines

  ‣ **autoencoders**

  ‣ sparse coding model

AUTOENCODER

Topics: autoencoder, encoder, decoder, tied weights

• Feed-forward neural network trained to reproduce its input at the output layer

$$\widehat{\mathbf{x}} \quad \mathbf{\widehat{x}}$$

Decoder

$$\widehat{\mathbf{x}} \ = \ o(\widehat{\mathbf{a}}(\mathbf{x}))$$

$$= \ \underbrace{\mathrm{sigm}}(\mathbf{c} + \mathbf{W}^*\mathbf{h}(\mathbf{x}))$$

for binary inputs

$$\mathbf{W}^* = \mathbf{W}^\top$$
(tied weights)

$$\mathbf{h}(\mathbf{x}) \quad b_j$$

$$\mathbf{W}$$

Encoder

$$\mathbf{h}(\mathbf{x}) \ = \ g(\mathbf{a}(\mathbf{x}))$$

$$= \ \mathrm{sigm}(\mathbf{b} + \mathbf{W}\mathbf{x})$$

$$\mathbf{x}$$

$$c_k$$