Neural networks

Autoencoder - denoising autoencoder
**OVERCOMPLET E HIDDEN LAYER**

**Topics:** overcomplete representation

- Hidden layer is overcomplete if greater than the input layer
  - no compression in hidden layer
  - each hidden unit could copy a different input component
- No guarantee that the hidden units will extract meaningful structure

\[
\hat{x} \quad c_k \\
\circ \quad \circ \quad \circ \\
\circ 
\]

\[
W^* = W^T \quad \text{(tied weights)}
\]

\[
h(x) \quad b_j \\
\circ \quad \circ \\
\circ 
\]

\[
x
\]
Topics: denoising autoencoder

- Idea: representation should be robust to introduction of noise:
  - random assignment of subset of inputs to 0, with probability $\nu$
  - Gaussian additive noise
- Reconstruction $\hat{x}$ computed from the corrupted input $\tilde{x}$
- Loss function compares $\hat{x}$ reconstruction with the noiseless input $x$
Topics: denoising autoencoder

• Idea: representation should be robust to introduction of noise:
  ‣ random assignment of subset of inputs to 0, with probability \( \nu \)
  ‣ Gaussian additive noise
• Reconstruction \( \hat{x} \) computed from the corrupted input \( \tilde{x} \)
• Loss function compares \( \hat{x} \) reconstruction with the noiseless input \( x \)
Topics: denoising autoencoder

- Idea: representation should be robust to introduction of noise:
  - random assignment of subset of inputs to 0, with probability \( \nu \)
  - Gaussian additive noise
- Reconstruction \( \hat{x} \) computed from the corrupted input \( \tilde{x} \)
- Loss function compares \( \hat{x} \) reconstruction with the noiseless input \( x \)
\[ \hat{x} = \text{sigm}(c + W^* h(\tilde{x})) \]
\[ p(\tilde{x}|x) \]
\[ \hat{x} = \text{sigm}(c + W^* h(\tilde{x})) \]

\[ p(\tilde{x}|x) \]
Denoising Autoencoder

\[ \hat{x} = \text{sigm}(c + W^* h(\tilde{x})) \]

\[ p(\tilde{x}|x) \]

\[ \text{Extracting and Composing Robust Features with Denoising Autoencoders} \]
\[ \hat{x} = \text{sigm}(c + W^* h(\tilde{x})) \]
FILTERS (DENOISING AUTOENCODER)
(Vincent, Larochelle, Bengio and Manzagol, ICML 2008)

- No corrupted inputs (cross-entropy loss)
FILTERS (DENOISING AUTOENCODER)
(Vincent, Larochelle, Bengio and Manzagol, ICML 2008)

- 25% corrupted inputs
FILTERS (DENOISING AUTOENCODER)
(Vincent, Larochelle, Bengio and Manzagol, ICML 2008)

- 50% corrupted inputs
SQUARED ERROR LOSS

- Training on natural image patches, with squared-difference loss
  - PCA is not the best solution
SQUARED ERROR LOSS

- Training on natural image patches, with squared-difference loss
  - Not equivalent to weight decay