Neural networks

Computer vision - convolutional RBM
Topics: convolutional RBM

- How about designing convolutional unsupervised networks
  - let’s consider the case of the RBM
  - could use same convolutional connectivity between input (v) and hidden layer (h)

\[
P(v, h) = \frac{1}{Z} \exp(-E(v, h))
\]

\[
E(v, h) = - \sum_{k=1}^{K} \sum_{i,j=1}^{N_H} \sum_{r,s=1}^{N_W} h_{ij}^k W_{rs}^k v_{i+r-1,j+s-1}
- \sum_{k=1}^{K} b_k \sum_{i,j=1}^{N_H} h_{ij}^k - c \sum_{i,j=1}^{N_V} v_{ij}.
= - \sum_{k=1}^{K} h_{ij}^k \cdot (\tilde{W}^k * v) - \sum_{k=1}^{K} b_k \sum_{i,j} h_{ij}^k - c \sum_{i,j} v_{ij}
\]

- \(h_{ij}^k\) are the hidden units of the \(k\)th feature map
- \(W_{rs}^k\) are the weights to the \(k\)th feature map
- \(\tilde{W}^k\) are the weights with flipped rows and columns (convolution kernel)

Lee et al. 2009
CONVOLUTIONAL RBM

**Topics:** convolutional RBM

- We can introduce a notion of probabilistic pooling
  - pooling unit $p^k_\alpha$ above is 1 only if at least one hidden unit $h^k_{i,j}$ in neighborhood is 1
  - within a pooling neighborhood, allow at most only a single unit $h^k_{i,j}$ equal to 1

---

**Lee et al. 2009**

$$I(h^k_{ij}) \triangleq b_k + (\tilde{W}^k \ast v)_{ij}$$

implies $p^k_\alpha$ is 1

$$P(h^k_{ij} = 1|v) = \frac{\exp(I(h^k_{ij}))}{1 + \sum_{(i',j') \in B_\alpha} \exp(I(h^k_{i',j'}))}$$

$$P(p^k_\alpha = 0|v) = \frac{1}{1 + \sum_{(i',j') \in B_\alpha} \exp(I(h^k_{i',j'}))}$$

implies all $h^k_{i,j}$ are 0
To illustrate the algorithm, we describe a case with one
the units of each layer are sampled in parallel (see Sec-
trivial solutions, such as feature detectors represent-
Our model is overcomplete in that the size of the rep-
3.4. Training via sparsity regularization
Convolutional Deep Belief Networks for Scalable Unsupervised Learning of Hierarchical Representations

\[ P(v_{ij} = 1|\mathbf{h}) = \sigma((\sum_k W^k \ast h^k)_{ij} + c) \]

\[ I(h^k_{ij}) \triangleq b_k + (\tilde{W}^k \ast v)_{ij} \]

imply \( P^k_{\alpha} \) is 1

\[ P(h^k_{ij} = 1|v) = \frac{\exp(I(h^k_{ij}))}{1 + \sum_{(i',j') \in B} \exp(I(h^k_{i',j'}))} \]

\[ P(p^k_{\alpha} = 0|v) = \frac{1}{1 + \sum_{(i',j') \in B} \exp(I(h^k_{i',j'}))} \]

imply all \( h^k_{i,j} \) are 0
CONVOLUTIONAL RBM

**Topics:** convolutional RBM

- Using these adapted conditionals, we can perform contrastive divergence
  - energy gradients involve convolutions, similar to the backprop gradients in convolutional network

- Can stack convolutional RBMs
  - provides a pretraining procedure which doesn’t require the extraction of patches

- See Lee et al. 2009 for more details

Lee et al. 2009