Neural networks

Sparse coding - dictionary update with projected gradient descent

SPARSE CODING

Topics: sparse coding

- For each $\mathbf{x}^{(t)}$ find a latent representation $\mathbf{h}^{(t)}$ such that:
  - it is sparse: the vector $\mathbf{h}^{(t)}$ has many zeros
  - we can reconstruct the original input $\mathbf{x}^{(t)}$ as much as possible
- More formally:

reconstruction error; sparsity penalty

$$\min_{\mathbf{D}} \frac{1}{T} \sum_{t=1}^{T} \min_{\mathbf{h}^{(t)}} \frac{1}{2} \overbrace{||\mathbf{x}^{(t)} - \underbrace{\mathbf{D}\ \mathbf{h}^{(t)}}_{}||_2^2} + \underbrace{\lambda}_{} \overbrace{||\mathbf{h}^{(t)}||_1}$$

reconstruction $\widehat{\mathbf{x}}^{(t)}$; reconstruction vs. sparsity control

  - $\mathbf{D}$ is equivalent to the autoencoder output weight matrix
  - however, $\mathbf{h}(\mathbf{x}^{(t)})$ is now a complicated function of $\mathbf{x}^{(t)}$
    - encoder is the minimization $\mathbf{h}(\mathbf{x}^{(t)}) = \arg\min_{\mathbf{h}^{(t)}} \frac{1}{2}||\mathbf{x}^{(t)} - \mathbf{D}\ \mathbf{h}^{(t)}||_2^2 + \lambda||\mathbf{h}^{(t)}||_1$

SPARSE CODING

Topics: dictionary update (algorithm 1)

- Going back to our original problem

$$\min_{\mathbf{D}} \frac{1}{T} \sum_{t=1}^{T} \min_{\mathbf{h}^{(t)}} l(\mathbf{x}^{(t)}) = \min_{\mathbf{D}} \frac{1}{T} \sum_{t=1}^{T} \frac{1}{2} ||\mathbf{x}^{(t)} - \mathbf{D}\, \mathbf{h}(\mathbf{x}^{(t)})||_2^2 + \lambda ||\mathbf{h}(\mathbf{x}^{(t)})||_1$$

- Let's assume $\mathbf{h}(\mathbf{x}^{(t)})$ doesn't depend on $\mathbf{D}$ (which is false)
  - we must minimize

$$\min_{\mathbf{D}} \frac{1}{T} \sum_{t=1}^{T} \frac{1}{2} ||\mathbf{x}^{(t)} - \mathbf{D}\, \mathbf{h}(\mathbf{x}^{(t)})||_2^2$$

  - we must also constrain the columns of $\mathbf{D}$ to be of unit norm

SPARSE CODING

Topics: dictionary update (algorithm 1)

- A gradient descent method could be used here too
  - specifically, this is a projected gradient descent algorithm
- While $\mathbf{D}$ hasn't converged
  - perform gradient update of $\mathbf{D}$

$$\mathbf{D} \Longleftarrow \mathbf{D} + \alpha \frac{1}{T} \sum_{t=1}^{T} (\mathbf{x}^{(t)} - \mathbf{D}\, \mathbf{h}(\mathbf{x}^{(t)}))\, \mathbf{h}(\mathbf{x}^{(t)})^{\top}$$

  - renormalize the columns of $\mathbf{D}$
    - for each column $\mathbf{D}_{\cdot,j}$:

$$\mathbf{D}_{\cdot,j} \Longleftarrow \frac{\mathbf{D}_{\cdot,j}}{||\mathbf{D}_{\cdot,j}||_2}$$