# Neural networks

Sparse coding - online dictionary learning algorithm

# SPARSE CODING

**Topics:** learning algorithm (putting it all together)

- Learning alternates between inference and dictionary learning

- While $\mathbf{D}$ has not converged
  - find the sparse codes $\mathbf{h}(\mathbf{x}^{(t)})$ for all $\mathbf{x}^{(t)}$ in my training set with ISTA
  - update the dictionary:
    - $\mathbf{A} \Longleftarrow \sum_{t=1}^{T} \mathbf{x}^{(t)}\, \mathbf{h}(\mathbf{x}^{(t)})^\top$
    - $\mathbf{B} \Longleftarrow \sum_{t=1}^{T} \mathbf{h}(\mathbf{x}^{(t)})\, \mathbf{h}(\mathbf{x}^{(t)})^\top$
    - run block-coordinate descent algorithm to update $\mathbf{D}$

- Similar to the EM algorithm

# SPARSE CODING

**Topics:** online learning algorithm

- This algorithm is "batch" (i.e. not online)
  - single update of the dictionary per pass on the training set
  - for large datasets, we'd like to update $\mathbf{D}$ after visiting each $\mathbf{x}^{(t)}$
- Solution: for each $\mathbf{x}^{(t)}$
  - perform inference of $\mathbf{h}(\mathbf{x}^{(t)})$ for the current $\mathbf{x}^{(t)}$
  - update running averages of the quantities required to update $\mathbf{D}$ :
    - $\mathbf{B} \Longleftarrow \beta\, \mathbf{B} + (1-\beta)\, \mathbf{x}^{(t)}\, \mathbf{h}(\mathbf{x}^{(t)})^\top$
    - $\mathbf{A} \Longleftarrow \beta\, \mathbf{A} + (1-\beta)\, \mathbf{h}(\mathbf{x}^{(t)})\, \mathbf{h}(\mathbf{x}^{(t)})^\top$
  - use current value of $\mathbf{D}$ as "warm start" to block-coordinate descent

# SPARSE CODING

**Topics:** online learning algorithm

- Initialize $\mathbf{D}$ (not to 0!)
- While $\mathbf{D}$ hasn't converged
  - for each $\mathbf{x}^{(t)}$
    - infer code $\mathbf{h}(\mathbf{x}^{(t)})$
    - update dictionary
      - ✓ $\mathbf{B} \Longleftarrow \beta\, \mathbf{B} + (1-\beta)\, \mathbf{x}^{(t)}\, \mathbf{h}(\mathbf{x}^{(t)})^\top$
      - ✓ $\mathbf{A} \Longleftarrow \beta\, \mathbf{A} + (1-\beta)\, \mathbf{h}(\mathbf{x}^{(T+1)})\, \mathbf{h}(\mathbf{x}^{(T+1)})^\top$
      - ✓ while $\mathbf{D}$ hasn't converged
        - ★ for each column $\mathbf{D}_{\cdot,j}$ perform gradient update

$$\mathbf{D}_{\cdot,j} \Longleftarrow \frac{1}{A_{j,j}} \left(\mathbf{B}_{\cdot,j} - \mathbf{D}\, \mathbf{A}_{\cdot,j} + \mathbf{D}_{\cdot,j}\, A_{j,j}\right)$$

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

Online Dictionary Learning for Sparse Coding.
Mairal, Bach, Ponce and Sapiro, 2009.