Neural networks
Sparse coding - online dictionary learning algorithm
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:
  - $\mathbf{A} \leftarrow \sum_{t=1}^{T} \mathbf{x}^{(t)} \mathbf{h}(\mathbf{x}^{(t)})^\top$
  - $\mathbf{B} \leftarrow \sum_{t=1}^{T} \mathbf{h}(\mathbf{x}^{(t)}) \mathbf{h}(\mathbf{x}^{(t)})^\top$
  - run block-coordinate descent algorithm to update $\mathbf{D}$

Similar to the EM algorithm
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 $D$ after visiting each $x(t)$
- Solution: for each $x(t)$
  - perform inference of $h(x(t))$ for the current $x(t)$
  - update running averages of the quantities required to update $D$:
    - $B \leftarrow \beta B + (1 - \beta) x(t) h(x(t))^T$
    - $A \leftarrow \beta A + (1 - \beta) h(x(t)) h(x(t))^T$
  - use current value of $D$ as “warm start” to block-coordinate descent
Topics: online learning algorithm

- Initialize $D$ (not to 0!)
- While $D$ hasn’t converged
  - for each $x(t)$
    - infer code $h(x(t))$
    - update dictionary
      - $B \leftarrow \beta B + (1 - \beta) x(t) h(x(t))^T$
      - $A \leftarrow \beta A + (1 - \beta) h(x^{(T+1)}) h(x^{(T+1)})^T$
      - while $D$ hasn’t converged
        - for each column $D_{.,j}$ perform gradient update
          - $D_{.,j} \leftarrow \frac{1}{A_{j,j}} (B_{.,j} - D A_{.,j} + D_{.,j} A_{j,j})$
          - $D_{.,j} \leftarrow \frac{D_{.,j}}{||D_{.,j}||_2}$