Neural networks
Sparse coding - dictionary learning algorithm
**Topics:** sparse coding

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

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

\( D \) is equivalent to the autoencoder output weight matrix

however, $h(x^{(t)})$ is now a complicated function of $x^{(t)}$

- encoder is the minimization $h(x^{(t)}) = \arg \min_{h^{(t)}} \frac{1}{2} \| x^{(t)} - D h^{(t)} \|_2^2 + \lambda \| h^{(t)} \|_1$
SPARSE CODING

Topics: learning algorithm (putting it all together)
- Learning alternates between inference and dictionary learning

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

• Similar to the EM algorithm