EE269
Signal Processing for Machine Learning
Lecture 20

Instructor : Mert Pilanci

Stanford University

November 18, 2020
Dictionary Learning

\[
\min_{x_1, \ldots, x_n, D} \sum_{i=1}^{n} \| Dx_i - y_i \|_2^2 \quad \text{s.t.} \quad \| x_i \|_0 \leq s
\]
Dictionary Learning

\[
\min_{x_1, \ldots, x_n, D} \sum_{i=1}^{n} \| Dx_i - y_i \|_2^2 \quad \text{s.t. } \| x_i \|_0 \leq s
\]

- Matrix factorization interpretation:
  \( DX \approx Y \) where \( X \) has sparse columns
Dictionary Learning: $\ell_1$ relaxation

$$\min_{x_1, \ldots, x_n, D} \sum_{i=1}^{n} \| Dx_i - y_i \|_2^2 + \lambda \| x_i \|_1$$
Dictionary Learning: \( \ell_1 \) relaxation

\[
\min_{x_1, \ldots, x_n, D} \sum_{i=1}^{n} \| Dx_i - y_i \|_2^2 + \lambda \| x_i \|_1
\]

- heuristic algorithm
- minimize over \( x_i \): \( \ell_1 \) regularized Least Squares problem
- gradient descent step over \( D \):
  \[
  D \leftarrow D - \mu \sum_i (Dx_i - y_i)x_i^T
  \]
Learned Dictionaries from Images

\[
\min_{x_1,...,x_n, D} \sum_{i=1}^{n} \|D x_i - y_i\|_2^2 + \lambda \|x_i\|_1
\]
Dictionary Learning vs Deep Learning

- Dictionary learning

- Alexnet: Convolutional net for image classification
Dictionary Learning Applications

[Mairal, Sapiro, and Elad, 2008d]
Since 1699, when French explorers landed at the great bend of the Mississippi River and celebrated the first Mardi Gras in North America, New Orleans has brewed a fascinating melange of cultures. It was French, then Spanish, then French again, then sold to the United States. Through all these years, and even into the 1900s, others arrived from everywhere: Acadians (Cajuns), Africans, indige-
Dictionary Learning Applications

Inpainting, [Mairal, Elad, and Sapiro, 2008b]
The dictionary can be structured, e.g., hierarchical, convolutional etc

\[
\min_{x_1,\ldots,x_n,D_1,D_2,\ldots,D_L} \sum_{i=1}^{n} \| D_1 D_2 \ldots D_L x_i - y_i \|_2^2 + \lambda \| x_i \|_1
\]
Clustering

$$\min \sum_{j=1}^{k} \sum_{x_i \in C_j} \| x_i - c_j \|^2 = \| X - CB^T \|_F^2$$

$C = [c_1, \ldots, c_k]$ is the centroid matrix

$B$ denotes the clustering assignment

$B_{ij} = 1$ if $i$-th observation is assigned to the $j$-th cluster center $c_j$
Matching Pursuit

Matching pursuit

\[
\min_{\alpha \in \mathbb{R}^m} \frac{1}{2} \| x - D\alpha \|_2^2 \quad \text{s.t.} \quad \| \alpha \|_0 \leq L
\]

1. Initialization: \( \alpha = 0 \), residual \( r = x \)
2. while \( \| \alpha \|_0 < L \)
3. Select the element with maximum correlation with the residual

\[ \hat{i} = \arg \max_{i=1,...,m} | d_i^T r | \]

4. Update the coefficients and residual

\[
\begin{align*}
\alpha_i &= \alpha_i + d_i^T r \\
r &= r - (d_i^T r) d_i
\end{align*}
\]

5. End while
Matching Pursuit: Example
Matching Pursuit: Example
Matching Pursuit: Example
Matching Pursuit: Example
Matching Pursuit: Example
Matching Pursuit: Example
Further applications

- Mechanical monitoring
  Engines, elevators, buildings, power plants etc.
- Medical monitoring
  Pulmonary (lung) sounds, cardiac sounds, speech etc.
- Security surveillance
  Interesting event discovery, accident detection etc.
Other SP/ML methods

- Independent Component Analysis
- Recurrent Neural Nets
- Generative Models and Generative Adversarial Networks
- Manifold learning methods (e.g., Laplacian Eigenmaps, Isomap)
- Tensor Decompositions (e.g. Tensor Principal Component Analysis)
- Latent and Mixture Models (Hidden Markov Models, Low rank models)
- Bayesian Methods
- State Space Models and tracking (e.g., Kalman filter)
Other Courses

- EE270 - Large Scale Matrix Computation, Optimization and Learning
- EE364a - Convex Optimization I
- EE364b - Convex Optimization II
- EE 264 - Digital Signal Processing
- EE373A - Adaptive Signal Processing
- CS 229 - Machine Learning
- CS 230 - Deep Learning