CS 6.5: Theories of Deep Learning
Problem Sheet 2

Prof. Jared Tanner

January 28, 2019

Submission Due: 10:00 AM, **05-02-2019, note one day ealier**

Label Consistent K-SVD:[1]
Label consistent K-SVD (LC-KSVD) is a discriminative dictionary learning algorithm based on the theory of sparse representations. It uses class labels of training data, to associate label information with each dictionary atom (columns of the dictionary matrix) to enforce discriminability in sparse codes. The overall objective of LC-KSVD is defined as:

$$< \mathbf{D}, \mathbf{A}, \mathbf{X}, \mathbf{W} >= \text{argmin} \, \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_2^2 + \alpha\|\mathbf{Q} - \mathbf{A}\mathbf{X}\|_2^2 + \beta\|\mathbf{H} - \mathbf{W}\mathbf{X}\|_2^2 \;\; \text{s.t.} \forall_i \;\; \|\mathbf{x}_i\|_0 \leq T, \tag{1}$$

where $\mathbf{D}$ is the dictionary, $\mathbf{Y}$ is the data matrix, $\mathbf{X}$ is the matrix of sparse codes, $\mathbf{H}$ is the label matrix, $\mathbf{A}$ is a transformation matrix, and $\mathbf{W}$ is the linear classification matrix.

The first term in eq(1) is to measure the reconstruction error, the second term enforces label consistency among sparse codes, and the third term learns a linear classifier. The solution to eq(1) is obtained using the K-SVD algorithm[2].

1. Task1: Write a short report summarizing the LC-KSVD algorithm, including the role of individual terms in eq(1) and their effect on the learned dictionary. Discuss how crucial are parameters α and β for the overall optimization. Note: there is no need to explain the K-SVD algorithm! just mention how it is employed. Your report should be written in the format and style of a NIPS Proceedings, abridged to not exceed 2 **(note, decreased from 3 to 2)** pages. Latex style files and an exemplar template are provided on the course page.

2. Task2: Let us assume that the algorithm is initialized with dictionary $\mathbf{D}^0$. Let $< \mathbf{D}^*, \mathbf{A}^*, \mathbf{W}^*, \mathbf{X}^* >$ be the solution of LC-KSVD for $\mathbf{Y} \in \mathbb{R}^{n \times N}, \mathbf{Q} \in \mathbb{R}^{K \times N}, \mathbf{H} \in \mathbb{R}^{m \times N}, \alpha, \beta$ and T, where n is the example dimension, N is the number of training examples, m is the number of classes, $K = mp$ is the dictionary size with p being the number of dictionary atoms learned per class.

   Prove that the solution $< \mathbf{D}^*, \mathbf{W}^*, \mathbf{X}^* >$ is also the solution for problem in eq(2) for $\mathbf{Y}, \mathbf{H}, T$ and $\gamma = p\alpha + \beta$

   $$< \mathbf{D}, \mathbf{A}, \mathbf{X} >= \text{argmin} \, \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_2^2 + \gamma\|\mathbf{H} - \mathbf{W}\mathbf{X}\|_2^2 \;\; \text{s.t.} \forall_i \;\; \|\mathbf{x}_i\|_0 \leq T \tag{2}$$

   given that $\tilde{\mathbf{A}} = \mathbf{P}_\pi \mathbf{A}$,

   $\tilde{\mathbf{A}}^* = \sqrt{\alpha}(\mathbf{W}^{*T} \ldots \mathbf{W}^{*T})$ ($\mathbf{W}$ repeated p times)

   $\mathbf{P}_\pi$ is a permutation matrix corresponding to purmutation π defined as: $\pi(i) = ((i-1) \bmod m)p + \lfloor \frac{i-1}{m} \rfloor + 1$.

   Hint: Try to find correspondence between $\mathbf{H}$ and $\mathbf{Q}$ and carefully go throught each and every step of the algorithm.

[1] LC-KSVD: http://users.umiacs.umd.edu/ zhuolin/Publications/0198.pdf
[2] K-SVD: https://elad.cs.technion.ac.il/wp-content/uploads/2018/02/32_KSVD_IEEE_TSP.pdf