Online Multi-Task Learning via Sparse Dictionary Optimization

Summary
We developed an efficient online method for learning multiple consecutive tasks based on the K-SVD algorithm for sparse dictionary optimization.

Capabilities of our ELLA-SVD algorithm:
- Learns multiple tasks concurrently
- Transfers knowledge to accelerate learning of new tasks
- Supports a variety of base learning algorithms
- Has lower computational cost than current lifelong learning algorithms
- Supports both task and feature similarity matrices

We demonstrate the effectiveness of ELLA-SVD in lifelong learning settings.

Introduction
Goal: Develop intelligent agents that
1. Quickly learn new tasks
2. Learn continually with experience
3. Exhibit versatility over multiple tasks

This work investigates a formulation of online multi-task learning (MTL) based on sparse dictionary optimization.

This approach builds upon our earlier work on the Efficient Lifelong Learning Algorithm (ELLA) [Eaton & Ruvolo, ICML ’13].

Background: Dictionary Learning for Sparse Coding via K-SVD
Goal: Given a data set \( \{x_1, \ldots, x_n\} \subset \mathbb{R}^d \), output a dictionary \( L \in \mathbb{R}^{d \times k} \) that sparse codes the data by solving:

\[
\arg\min_{L} \sum_{i=1}^{n} \min_{s_i} \left\{ \|Ls_i - x_i\|_2^2 + \mu \|s_i\|_0 \right\}
\]

The K-SVD Algorithm
Iterate two steps until convergence to yield \( L \):

Step 1: update codes for each point
\( s_i^{(t)} \leftarrow \arg\min_{s_i} \left\{ \|Ls_i - x_i\|_2^2 + \mu \|s_i\|_0 \right\} \)

Step 2: update each basis vector and the weights of the data points that utilize this basis vector
\( m \in A \Leftrightarrow s_i^{(m)} \neq 0 \)
\( l_j, s_j^{(t)} \leftarrow \arg\min_{l_j, s_j} \sum_{i=1}^{n} \left\{ \|Ls_i^{(t)} - x_i\|_2^2 + \mu \|s_i^{(t)}\|_0 \right\} \)

Step 2 can be solved efficiently via SVD:
- Let the \( m \)th column of \( E \) be given by \( e_j = x_j - \sum_{r \neq j} s_j^{(t)} l_r \)

- Then take

\[
(U, \Sigma, V) = \text{svd} (E_A)
\]

\( l_j \leftarrow u_j \quad s_j^{(t)} \leftarrow \sigma_{1,j} v_1 \)

Surprisingly, we can efficiently find the global minimum!

Online Multi-Task Learning via K-SVD

1. Tasks are received sequentially
2. Knowledge is transferred from previously learned tasks
3. New knowledge is stored for future use
4. Existing knowledge is refined

Lifelong Learning System

ELLA-SVD Algorithm
Given a new task \( t \),
1. Train a single-task model \( \theta^{(t)} \) for task \( t \)
2. Reconstruct \( \theta^{(t)} \) in the current basis (LASSO):
\[ s(t) = \arg\min_{s} \left\{ \|Ls - \theta^{(t)}\|_2^2 + \mu \|s\|_1 \right\} \]
3. Update the basis:
\[ s_j^{(t)} \leftarrow \arg\min_{s_j} \sum_{i=1}^{n} \|s_j\|_0 \]

Per-Task Computational Complexity
ELLA-SVD: \( O(\text{base learner} + \text{d} \cdot k^2 \cdot d + q \cdot d^2) \)

Applications
Facial Expression Recognition: identify presence of facial action units (RS4 upper lid raiser, #10 upper lip raiser, #12 lip corner pull)

Land Mine Detection from radar

Student Exam Score Prediction

ELLA Incremental – a more efficient but suboptimal version of ELLA

ELLA Dual Update – a hybrid combination of ELLA-SVD & ELLA Incremental

Results
We compared ELLA-SVD to ELLA and two variants:
- ELLA Incremental – a more efficient but suboptimal version of ELLA
- ELLA Dual Update – a hybrid combination of ELLA-SVD & ELLA Incremental

Acknowledgement: This research was supported by ONR grant N00014-11-1-0139

Sparse dictionary optimization provides a computationally efficient method for online multi-task learning.