Label-consistent sparse auto-encoders

Thomas Rolland  
IRIT, UPS, CNRS, Toulouse, France  
INESC-ID, Lisbon, Portugal  
Email: thomas.rolland91@orange.fr

Adrian Basarab  
IRIT, Université Paul Sabatier, CNRS  
Toulouse, France  
Email: adrian.basarab@irit.fr

Thomas Pellegrini  
IRIT, Université Paul Sabatier, CNRS  
Toulouse, France  
Email: thomas.pellegrini@irit.fr

Auto-encoders (AE) are a particular type of unsupervised neural networks that aim at providing a compact representation of a signal or an image [1]. Such AEs are useful for data compression but most of the time the representations they provide are not appropriate as is for a downstream classification task. This is due to the fact that they are trained to minimize a reconstruction error and not a classification loss. Classification attempts with AEs have already been proposed such as contractive AEs [2], correspondence AEs [3] and stacked similarity-aware AEs [4], for instance. Inspired by label-consistent K-SVD (LC-KSVD) [5], we propose a novel supervised version of AEs that integrates class information within the encoded representations.

I. LABEL-CONSISTENT SPARSE CODING (LC-KSVD)

Sparse Coding (SC) and (sparse) AEs share a similar objective of providing compact data representations. LC-KSVD consists in adding to the standard SC reconstruction error objective: i) a label consistency constraint (a "discriminative sparse-code error"), ii) a classification error term. It results in a unified objective function that can be solved with the standard K-SVD algorithm. To do this, Jiang et al [5] defined the following objective function:

\[
(D, A, W, \gamma) = \arg \min_{D, A, W, \gamma} \|X - D\gamma\|^2_0 + \mu\|Q - A\gamma\|^2_2 + \beta\|H - W\gamma\|^2_2
\]

where \(D\) and \(\gamma\) are respectively the dictionary and sparse codes to be estimated, \(Q\) is a matrix of discriminative sparse codes of the input signals \(X\). \(A\) a linear transformation matrix, \(W\) a linear classifier and \(H\) the labels associated to \(X\). \(Q\) arbitrarily associates to an input signal a number of dictionary atoms, with non-zero values occurring when signal \(i\) and atom \(k\) share the same label (see Fig 1). \(Q\) is arbitrarily defined by the user with the possibility to let some atoms "empty" by not assigning them any class (in white in Fig. 1).

II. PROPOSED LABEL-CONSISTENT AUTOENCODERS (LC-Æ)

Fig. 2 shows the architecture of the proposed LC-Æ comprised of a standard sparse convolutional AE central part, completed with a "Q branch" and an "H branch" to emulate the label-consistent and the consistency terms from (1). These branches are fully-connected layers with sigmoid and softmax activation layers for Q and H, respectively. The AE was trained with the cross-entropy cost function: the binary variant for the reconstruction objective and the Q-branch output, and the categorical variant for the classification H-branch.

III. EXPERIMENTS

We compare the feature representation methods on MNIST. After extracting the sparse representations with each method, we train and test k-means and SVM with radial kernel (RBF) on the training and evaluation subsets of MNIST comprised of 50k and 10k images, respectively. SVM and k-means allow to compare the discriminative power of the representations in supervised and unsupervised settings.

The hyperparameters were tuned for classification. For the sparse coding approaches (standard SC and LC-KSVD), we used 1024 (about twice the dimension of the images \(d = 728\)) atoms for the dictionaries and \(\lambda = 1.2/\sqrt{728}\) as suggested in [6]. For LC-KSVD, we used \(\mu = 5.0\) and \(\beta = 2.0\), which are large values to promote discriminative power over reconstruction [5].

For the proposed LC-Æs, 16 atoms per class (160 in total) gave satisfactory results as reported here-after. The encoder part is comprised of three \(3 \times 3\) convolution layers (16-10-10 filters respectively) with a rectifier-linear unit activation function, each followed by a \(2 \times 2\) max-pooling layer for sub-sampling. The encoder output representations are 160-d vectors. Two variants of the proposed model are compared to a standard sparse AE (denoted AE hereafter): a sparse AE with the "Q-branch" (LC-Æ-I), a sparse AE with the "H-branch" (LC-Æ-II). The \(\ell_1\)-norm sparse regularization coefficient was tuned to \(1 \leq 7\).

Fig. 3 shows examples of nine digit images from the MNIST eval subset with the original, reconstructed images on the first and second rows. The third and fourth rows show respectively the sparse codes and the discriminative representations obtained with an LC AE. These correspond to vector outputted by the AE that we reshaped in 2-d images for illustration. As can be seen, the reconstructed images are close to their original counterparts. Regarding the encoded activations shown in the third row, one can identify patterns similar between two instances of a same digit, for instance digit "1". Finally, the bottom row shows the outputs of the Q branch that were reshaped to have 10 columns corresponding to 16 atoms for each class as was defined the Q matrix. Learning was successful since one can most often very clearly identify each digit based on the column that is the darkest.

Table I gives a performance comparison between the different methods when using k-means (purity) and SVM (accuracy). For AEs, we always score the representation outputted by the encoder part of the model. As can be seen, SC and the sparse AE are not successful in providing discriminative representations that work with both clustering and SVM since purity values are close to chance (10%). Adding label-consistency constraints, either to SC or AEs, drastically improve the representation separability, with 78% purity for LC-KSVD, and 96-97% purity for LC-Æs with one and two branches. Finally, the LC-Æs give the best results with the SVM classifier, showing that with only 16 atoms per class instead of about 100 with LC-KSVD, these models provide very discriminative encoded representations.

We showed in this work that the proposed LC-Æs are effective in providing representations that allow for satisfactory image reconstructions and that embed discriminative information about the image classes. Ongoing experiments on other datasets are being conducted, such as tiny-imagenet and sound recordings (ESC-10), and similar trends are observed.
Fig. 1. Example for the user-defined Q matrix, each color corresponds to a class. In this example, signals 1, 3 and 6 belong to class 1; signals 2 and 5 to class 2 and signal 5 to class 3. Atom k8 is unassigned.

<table>
<thead>
<tr>
<th>Atom</th>
<th>Signal 1</th>
<th>Signal 2</th>
<th>Signal 3</th>
<th>Signal 4</th>
<th>Signal 5</th>
<th>Signal 6</th>
</tr>
</thead>
<tbody>
<tr>
<td>k1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>k2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>k3</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>k5</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>k6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>k7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>k8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Fig. 2. The proposed LC-AE architecture.

Fig. 3. MNIST samples and representations obtained with a Sparse LC-AE: original images (top row), reconstructed images (second-top row), encoded representations (third row), "Q-branch" representations (bottom row).

REFERENCES


TABLE I

<table>
<thead>
<tr>
<th>Method</th>
<th>k-means</th>
<th>SVM (RBF)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sparse Coding</td>
<td>0.13</td>
<td>0.90</td>
</tr>
<tr>
<td>LC-KSVD</td>
<td>0.78</td>
<td>0.91</td>
</tr>
<tr>
<td>AE</td>
<td>0.19</td>
<td>0.96</td>
</tr>
<tr>
<td>LC-AE-1</td>
<td>0.96</td>
<td>0.98</td>
</tr>
<tr>
<td>LC-AE-2</td>
<td>0.97</td>
<td>0.99</td>
</tr>
</tbody>
</table>

PERFORMANCE COMPARISON IN TERMS OF PURITY FOR K-MEANS AND ACCURACY FOR SVM.