The Importance of Encoding Versus Training with Sparse Coding and Vector Quantization[1]

Yingzhen Li

Department of Engineering

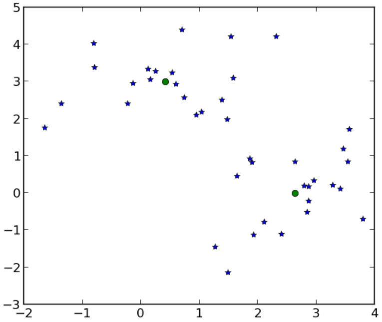
Feb. 2014

[1] See [1] for details.

Vector Quantization [2]

- Data $\xrightarrow{Quantizer}$ Features $\xrightarrow{Classifier}$ Predictions

$$\min_{C_k, W} \sum_k \sum_{\mathbf{x}^{(i)} \in C_k} ||W_k - \mathbf{x}^{(i)}||_2^2$$

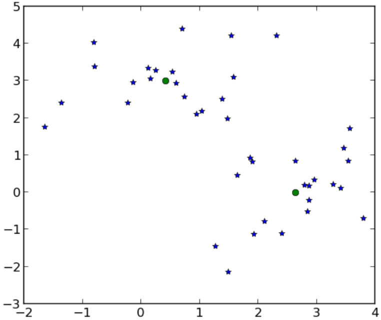


Dictionary Learning

- Data $\xrightarrow{\textit{Encoding}}$ Features $\xrightarrow{\textit{Classifier}}$ Predictions

$$\min_{\mathbf{s}^{(i)}, D} \sum_i ||D\mathbf{s}^{(i)} - \mathbf{x}^{(i)}||_2^2$$

- Orthogonal matching pursuit (OMP-k): restrict $||\mathbf{s}^{(i)}||_0 \leq k, \forall i$
- Sparse coding: add sparsity penalty (e.g. l_1-penalty)



Dictionary Construction

- Sparse Coding (SC)

$$\min_{\mathbf{s}^{(i)}, D} \sum_i ||D\mathbf{s}^{(i)} - \mathbf{x}^{(i)}||_2^2 + \lambda ||\mathbf{s}^{(i)}||_1 \tag{1}$$

- Orthogonal Matching Pursuit (OMP-k)

$$\min_{\mathbf{s}^{(i)}, D} \sum_i ||D\mathbf{s}^{(i)} - \mathbf{x}^{(i)}||_2^2 \quad s.t. \ ||\mathbf{s}^{(i)}||_0 \leq k, \forall i \tag{2}$$

- Sparse RBMs and Sparse Auto-encoders (SAEs)

$$P(\mathbf{s}^{(i)} | D, \mathbf{x}^{(i)}) \propto sigmoid(D^T \mathbf{x}^{(i)}) \tag{3}$$

- Randomly Sampled Patches (RP)
- Random Weights

Feature Mapping (with D fixed)

- Sparse Coding (SC)
  - for new data point $\mathbf{x}$, compute the code $\mathbf{s}$ by minimizing (1)
  - define feature $\mathbf{f} = [\max(\mathbf{0}, \mathbf{s}), \max(\mathbf{0}, -\mathbf{s})]$
- OMP-k: as above (by solving (2))
  - $\mathbf{f} = [\max(\mathbf{0}, \mathbf{s}), \max(\mathbf{0}, -\mathbf{s})]$
- Soft Threshold (T)
  - $\mathbf{f} = [\max(\mathbf{0}, D^T\mathbf{x} - \alpha), \max(\mathbf{0}, -D^T\mathbf{s} - \alpha)]$
- 'Natural' Encoding
  - $\mathbf{f} = [sigmoid(D^T\mathbf{x} + \mathbf{b}), sigmoid(-D^T\mathbf{s} + \mathbf{b})]$

Results (with L2-SVM)

Dict \ Encode	Natural	SC	OMP-1	OMP-10	T
R	70.5	74.0	65.8	68.6	73.2
RP	76.0	76.6	70.1	71.6	78.1
RBM	74.1	76.7	69.5	72.9	78.3
SAE	74.8	76.5	68.8	71.5	76.7
SC	77.9	78.5	70.8	75.3	78.5
OMP-1	71.4	78.7	71.4	76.0	78.9
OMP-2	73.8	78.5	71.0	75.8	79.0
OMP-5	75.4	78.8	71.0	76.1	79.1
OMP-10	75.3	79.0	70.7	75.3	79.4

Table: Cross-validation results (accuracy in %) on CIFAR-10.

Results (with L2-SVM)

Dict \ Encode	Natural	SC ($\lambda = 1$)	T ($\alpha = 0.5$)
R	91.9	93.8	93.1
RP	92.8	95.0	93.6
SC ($\lambda = 1$)	94.1	94.1	93.5
OMP-1	90.9	94.2	92.6

Table: Test accuracies (%) for the NORB jittered-cluttered dataset.

Results (with L2-SVM)

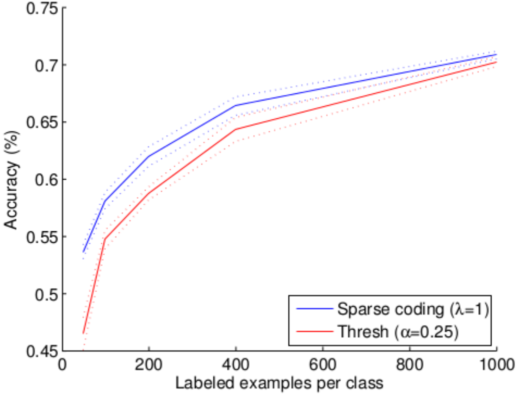


Figure: Performance of sparse coding and soft-threshold activations on CIFAR-10.

Conclusion

- Sparse coding as the encoder
  - Virtually any training algorithm can be used to create a suitable dictionary.
  - Yielding better representations with small training set.
- Encoder with dictionary can often be competitive with sparse coding.
  - VQ: ensure the tracking of local features

Adam Coates and Andrew Y. Ng.
The importance of encoding versus training with sparse coding and vector quantization.
In *ICML*, 2011.

Ernest Mwebaze, Petra Schneider, Frank-Michael Schleif, Sven Haase, Thomas Villmann, and Michael Biehl.
Divergence based learning vector quantization.
In *ESANN*, 2010.