Atom extraction and dictionary learning

Dictionary learning is a technique which allows rebuilding a sample starting from a sparse dictionary of atoms (similar to principal components). In Mairal J., Bach F., Ponce J., Sapiro G., **Online Dictionary Learning for Sparse Coding**, Proceedings of the 29th International Conference on Machine Learning, 2009 there's a description of the same online strategy adopted by scikit-learn, which can be summarized as a double optimization problem where:

$$X = \{\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_n\} \, where \, \bar{x}_i \in \mathbb{R}^m$$

Is an input dataset and the target is to find both a dictionary **D** and a set of weights for each sample:

$$\boldsymbol{D} \in \mathbb{R}^{m \times k} \, and \, \mathrm{A} = \{\bar{\alpha}_1, \bar{\alpha}_2, \ldots, \bar{\alpha}_m\} \, where \, \bar{\alpha}_i \in \mathbb{R}^k$$

After the training process, an input vector can be computed as:

$$\bar{x}_i = \boldsymbol{D}\bar{\alpha}_i$$

The optimization problem (which involves both **D** and alpha vectors) can be expressed as the minimization of the following loss function:

$$L(\boldsymbol{D}, A) = \frac{1}{2}\sum_{i} \|x_i - D\bar{\alpha}_i\|_2^2 + c\,\|\bar{\alpha}_i\|_1$$

Here the parameter **c** controls the level of sparsity (which is proportional to the strength of **L1** normalization). This problem can be solved by alternating the least square variable until a stable point is reached...

Previous Section (/book/big_data_and_business_intelligence/9781785889622/3/ch03lvl1s

Next Section (/book/big_data_and_business_intelligence/9781785889622/3/ch03lvl1sec2