Partitioning of Image Datasets using Discriminative Context Information

Christoph H. Lampert
Max-Planck-Institute for Biological Cybernetics, Tübingen, Germany

Overview

Summary
- Discriminative Context Partitioning (DCP) is a new unsupervised method to partition a dataset.
- It splits the dataset such that the resulting parts are best separated from a disjoint context class.
- DCP is not clustering. The parts are not determined by peaks in the sample density, but purely discriminatively.
- For suitable context, DCP is more robust than clustering methods.
- By varying the context, one can explore different partitions.

How to split a unimodal dataset into meaningful parts?

Use separation from a geometric context to distinguish parts

Method

A measure of separation between sets
For two disjoint sets \( X, Z \) and a decision hyperplane \( f \in \mathcal{H} \), we measure the separation between the sets by the negative of the SVM objective function:

\[
\text{sep}(X, Z) := -\frac{1}{2} \| f \|_2^2 - \sum_{x \in X} (1 - f(x)) - \sum_{z \in Z} (1 + f(z)).
\]

where \( f \) is a monotonous convex function that penalizes margin violations, e.g. the hinge loss or the quadratic loss.

The most discriminative split of a sets
Let \( X \) be the dataset we want to split. Let \( Z \) be a disjoint context set. For \( K \in \mathbb{N} \), let \( X_1 \cup \cdots \cup X_K = X \) be a decomposition of \( X \). Then the total separation score of this split is:

\[
\text{sep}(X_1, \ldots, X_K, Z) := \sum_{k=1}^{K} \max_{f_k \in \mathcal{H}} \text{sep}(X_k, Z).
\]

A decomposition \( X_1 \cup \cdots \cup X_K = X \) is called a most discriminative \( K \)-split of \( X \) with respect to \( Z \), if it maximizes the total separation over all possible decompositions of \( X \).

Theorem: Finding the most discriminative partitioning
The most discriminative partitioning of \( X \) with respect to \( Z \) is given by

\[
X_k^* := \{ x \in X : \arg \max f_k^*(x) = k \},
\]

for \( k = 1, \ldots, K \), where \( f_k^* \in \mathcal{H} \) minimizes

\[
J(t_1, \ldots, t_K) = \sum_{k=1}^{K} \| A_k \|_{1, 2}^2 + \sum_{k=1}^{K} (1 + t_k) + \sum_{k=1}^{K} (1 - t_k).
\]

Numeric Solution
Several techniques are applicable to solve the optimization problem (3):
- (Stochastic) gradient descent
- Convex-Concave Procedure (CCCP)
- Deterministic Annealing

Experimental Results
Unsupervised separation of USPS digits
We create an image dataset by mixing two USPS digit classes in different ratios. How well can we recover the original partitioning?

Finding substructures within Caltech256 classes
By varying the context class, we can browse through different splits of a single-label dataset. To humans, such splits can be interpretable:

Figure: Explorative use of DCP, the brain class in Caltech256 is split using two different context classes. A human could interpret the first split as structured vs. smooth, and the second as natural vs. schematic.