Pattern Recognition (PR)

Prof. Dr.-Ing. Andreas Maier
Pattern Recognition Lab (CS 5), Friedrich-Alexander-Universität Erlangen-Nürnberg
Winter Term 2020/21

This is a printable version of the slides of the lecture

Pattern Recognition (PR)
Winter term 2020/21
Friedrich-Alexander University of Erlangen-Nuremberg.

Erlangen, January 8, 2021
Prof. Dr.-Ing. Andreas Maier

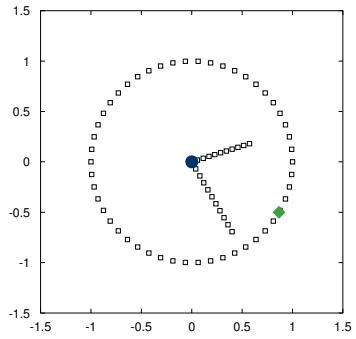
Laplacian Support Vector Machines

Pattern Recognition Lab

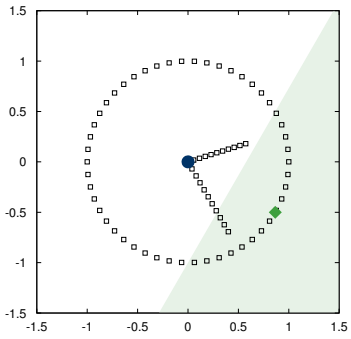
FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
FACULTY OF ENGINEERING

Motivation

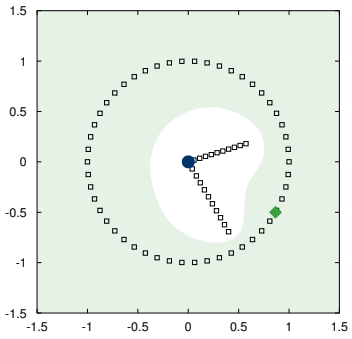
Example: two class “clock” data set



(a) Large set of unlabeled samples (black squares) and only one labeled sample per class (blue circle, green diamond)



(b) Result of a maximum margin supervised classification



(c) Result of a semi-supervised classification with intrinsic norm from manifold regularization

Learning from labeled and unlabeled data

Training data: $\mathscr{S} = \mathscr{L} \cup \mathscr{U}$

- labeled data: $\mathscr{L} = \{(\boldsymbol{x}_i, y_i), \quad i = 1, \ldots, l\}$
- unlabeled data: $\mathscr{U} = \{\boldsymbol{x}_i, \quad i = \underbrace{l+1, \ldots, m}_{u}\}$

Graph Laplacian $\boldsymbol{L}$ associated with $\mathscr{S}$:

- $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{W}$
- adjacency matrix $\boldsymbol{W}$
- diagonal matrix $\boldsymbol{D}$ with the degree of each node: $d_{ii} = \sum_{j=1}^{m} w_{ij}$

Kernel matrix $\boldsymbol{K}$: $k_{ij} = k(\boldsymbol{x}_i, \boldsymbol{x}_j)$

Decision boundary $f(\boldsymbol{x})$: $\boldsymbol{f} = [f(\boldsymbol{x}_i), \; i = 1, \ldots, m]^T$

Learning from labeled and unlabeled data (cont.)

Regularization framework for function learning:

$$f^* = \underset{f \in \mathscr{H}_k}{\operatorname{argmin}} \sum_{i=1}^{l} V(\boldsymbol{x}_i, y_i, f) + \gamma_A \|f\|_A^2 + \gamma_I \|f\|_I^2$$

Loss function $V(\boldsymbol{x}_i, y_i, f)$

- Squared loss function $(y_i - f(\boldsymbol{x}_i))^2$ for Regularized Least Squares (RLS)
- Hinge loss function $\max[0, 1 - y_i f(\boldsymbol{x}_i)]$ for SVM

Learning from labeled and unlabeled data (cont.)

Regularization framework for function learning:

$$f^* = \underset{f \in \mathcal{H}_k}{\operatorname{argmin}} \sum_{i=1}^{l} V(\boldsymbol{x}_i, y_i, f) + \gamma_A \|f\|_A^2 + \gamma_I \|f\|_I^2$$

Regularization terms

- *Ambient* norm $\|\cdot\|_A$:
  - norm of the function f in the Reproducing Kernel Hilbert Space (RKHS)
  - enforces a smoothness condition on the possible solutions
- *Intrinsic* norm $\|\cdot\|_I$:
  - norm of the function f in the low dimensional manifold
  - enforces a smoothness along the sampled $\mathcal{M}$

Reproducing Kernel Hilbert Spaces (RKHS)

Hilbert space

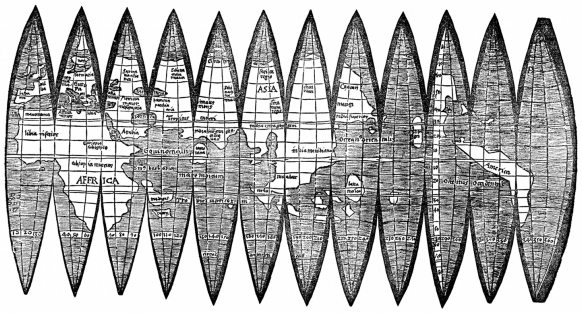
- abstract vector space with any finite or infinite number of dimensions
- possesses the structure of the inner product
- allows the measurement of angles and lengths
- is complete

Reproducing Kernel Hilbert Space

- Hilbert space of functions
- can be defined by kernels

Manifolds

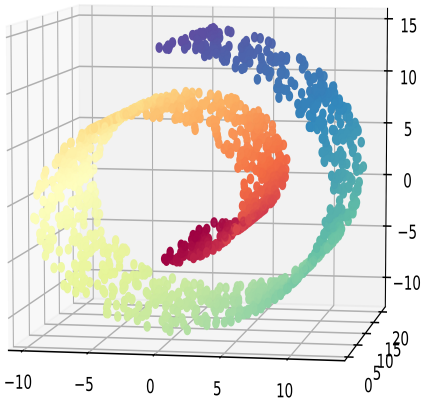
A manifold is a topological space that on a small enough scale resembles the Euclidean space.

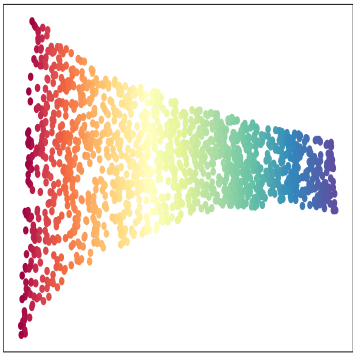


Martin Waldseemüller, Public domain, via Wikimedia Commons

Manifold Learning

The Swiss Roll Problem





Algorithm: S. Marsland, "Machine Learning: An Algorithmic Perspective", Chapter 10, 2009.

Learning from labeled and unlabeled data (cont.)

Intrinsic norm $\|\cdot\|_I$:

$$\|f\|_I^2 = \sum_{i=1}^{m}\sum_{j=i}^{m} w_{ij}\big(f(\boldsymbol{x}_i) - f(\boldsymbol{x}_j)\big)^2 = \boldsymbol{f}^T \boldsymbol{L}\, \boldsymbol{f}$$

Pattern Recognition Lab
FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG FACULTY OF ENGINEERING

Learning from labeled and unlabeled data (cont.)

Representer Theorem (Kimeldorf and Wahba, 1970):

The solution f^* of this optimization problem has the form:

$$f^*(\boldsymbol{x}) = \sum_{i=1}^{m} \beta_i^* \cdot k(\boldsymbol{x}_i, \boldsymbol{x}) + \beta_0^*$$

Laplacian Support Vector Machines

Constrained primal optimization problem based on the dual form:

$$\min_{\boldsymbol{\beta}\in\mathbb{R}^m,\boldsymbol{\xi}\in\mathbb{R}^l} \quad \sum_{i=1}^{l}\xi_i + \gamma_A\cdot\boldsymbol{\beta}^T\boldsymbol{K}\boldsymbol{\beta} + \gamma_I\cdot\boldsymbol{\beta}^T\boldsymbol{K}\boldsymbol{L}\boldsymbol{K}\boldsymbol{\beta}$$

$$\text{subject to} \quad y_i\left(\sum_{j=1}^{m}\beta_j k(\boldsymbol{x}_j,\boldsymbol{x}_i)+\beta_0\right)\geq 1-\xi_i, \quad i=1,\ldots,l$$

$$\xi_i\geq 0, \quad i=1,\ldots,l$$

Laplacian Support Vector Machines (cont.)

Lagrange function L:

$$\begin{aligned}L(\boldsymbol{\beta},\beta_0,\boldsymbol{\xi},\boldsymbol{\lambda},\boldsymbol{\nu}) \quad = \quad & \sum_{i=1}^{l}\xi_i+\frac{1}{2}\boldsymbol{\beta}^T(2\gamma_A\boldsymbol{K}+2\gamma_I\boldsymbol{K}\boldsymbol{L}\boldsymbol{K})\boldsymbol{\beta}- \\ & -\sum_{i=1}^{l}\lambda_i\left(y_i\left(\sum_{j=1}^{m}\beta_i k(\boldsymbol{x}_i,\boldsymbol{x}_j)+\beta_0\right)-1+\xi_i\right)- \\ & -\sum_{i=1}^{l}\nu_i\xi_i\end{aligned}$$

Laplacian Support Vector Machines (cont.)

KKT condition: the gradient w. r. t. the primal variables $\boldsymbol{\beta}, \beta_0, \boldsymbol{\xi}$ has to vanish

- Partial derivative w. r. t. β_0:

$$\frac{\partial L}{\partial \beta_0} \stackrel{!}{=} 0 \quad \Rightarrow \quad \sum_{i=1}^{I} \lambda_i y_i = 0$$

- Partial derivative w. r. t. ξ_i:

$$\frac{\partial L}{\partial \xi_i} \stackrel{!}{=} 0 \quad \Rightarrow \quad 1 - \lambda_i - \nu_i = 0 \quad \Rightarrow \quad 0 \leq \lambda_i \leq 1$$

Laplacian Support Vector Machines (cont.)

Simplifying the Langrangian using the two identities above:

$$\begin{aligned} L(\boldsymbol{\beta}, \boldsymbol{\lambda}) &= \frac{1}{2}\boldsymbol{\beta}^T(2\gamma_A \boldsymbol{K} + 2\gamma_I \boldsymbol{K}\boldsymbol{L}\boldsymbol{K})\boldsymbol{\beta} - \\ &\quad - \sum_{i=1}^{I} \lambda_i \left(y_i \left(\sum_{j=1}^{m} \beta_i k(\boldsymbol{x}_i, \boldsymbol{x}_j) + \beta_0 \right) - 1 \right) \\ &= \frac{1}{2}\boldsymbol{\beta}^T(2\gamma_A \boldsymbol{K} + 2\gamma_I \boldsymbol{K}\boldsymbol{L}\boldsymbol{K})\boldsymbol{\beta} - \boldsymbol{\beta}^T \boldsymbol{K}\boldsymbol{J}_{\mathscr{L}}^T \boldsymbol{Y}\boldsymbol{\lambda} + \sum_{i=1}^{I} \lambda_i \end{aligned}$$

with $\boldsymbol{J}_{\mathscr{L}} = [\boldsymbol{I}\ 0] \in \mathbb{R}^{I \times m}$:

- identity matrix $\boldsymbol{I} \in \mathbb{R}^{I \times I}$
- rectangular matrix $0 \in \mathbb{R}^{I \times u}$ with all entries being 0

and diagonal matrix $\boldsymbol{Y} \in \mathbb{R}^{I \times I}$ composed by the I class labels y_i

Laplacian Support Vector Machines (cont.)

Partial derivative w. r. t. $\boldsymbol{\beta}$:

$$\frac{\partial L}{\boldsymbol{\beta}} \stackrel{!}{=} 0 \quad \Rightarrow \quad (2\gamma_A \boldsymbol{K} + 2\gamma_I \boldsymbol{K}\boldsymbol{L}\boldsymbol{K})\boldsymbol{\beta} - \boldsymbol{K}\boldsymbol{J}_{\mathscr{L}}^T \boldsymbol{Y}\boldsymbol{\lambda} = 0$$

$$\Rightarrow \quad \boldsymbol{\beta} = (2\gamma_A \boldsymbol{I} + 2\gamma_I \boldsymbol{K}\boldsymbol{L})^{-1} \boldsymbol{J}_{\mathscr{L}}^T \boldsymbol{Y}\boldsymbol{\lambda}$$

Note: direct relationship between parameters $\boldsymbol{\beta}$ and Lagrange multipliers $\boldsymbol{\lambda}$

Laplacian Support Vector Machines (cont.)

Substituting back in the Langrange expression leads to the dual problem:

$$\begin{aligned} \max_{\boldsymbol{\lambda} \in \mathbb{R}^l} \quad & \sum_{i=1}^{l} \lambda_i - \frac{1}{2} \boldsymbol{\lambda}^T \boldsymbol{Q} \boldsymbol{\lambda} \\ \text{subject to} \quad & 0 \leq \lambda_i \leq 1, \quad i = 1, \ldots, l \\ & \sum_{i=1}^{l} \lambda_i y_i = 0 \end{aligned}$$

where

$$\boldsymbol{Q} = \boldsymbol{Y}\boldsymbol{J}_{\mathscr{L}}\boldsymbol{K}(2\gamma_A \boldsymbol{I} + 2\gamma_I \boldsymbol{K}\boldsymbol{L})^{-1} \boldsymbol{J}_{\mathscr{L}}^T \boldsymbol{Y}$$

Lessons Learned

Laplacian SVM:

- Ongoing research topic
- Extension of the Kernel SVM
- Additional regularization term
- Derivation of the dual problem

Pattern Recognition Lab

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
TECHNISCHE FAKULTÄT

Next Time in Pattern Recognition



Further Readings

- Stefano Melacci, Mikhail Belkin:
  Laplacian Support Vector Machines Trained in the Primal,
  Journal of Machine Learning Research, 12:1149-1184, 2011
- Mikhail Belkin, Partha Niyogi, Vikas Sindhwani:
  Manifold Regularization: A Geometric Framework for Learning from Labeled and Unlabeled Examples,
  Journal of Machine Learning Research, 7:2399-2434, 2006