

These are the slides of the lecture

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

These slides are released under Creative Commons License Attribution CC BY 4.0.

Please feel free to reuse any of the figures and slides, as long as you keep a reference to the source of these slides at <https://lme.tf.fau.de/teaching/> acknowledging the authors Niemann, Hornegger, Hahn, Steidl, Nöth, Seitz, Rodriguez, Das and Maier.

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

Pattern Recognition (PR)

Prof. Dr.-Ing. Andreas Maier

Pattern Recognition Lab (CS 5), Friedrich-Alexander-Universität Erlangen-Nürnberg
Winter Term 2020/21

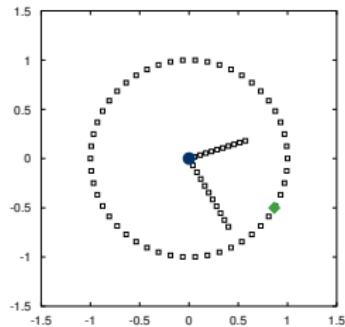


Laplacian Support Vector Machines

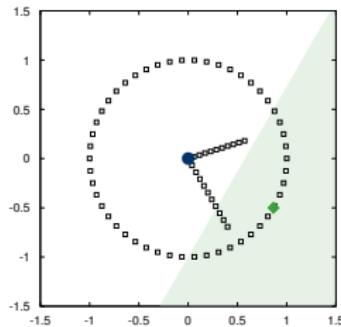


Motivation

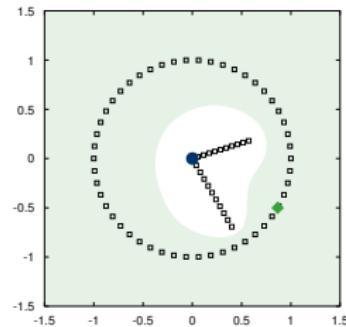
Example: two class “clock” data set



(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: $\mathcal{S} = \mathcal{L} \cup \mathcal{U}$

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

Learning from labeled and unlabeled data

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

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

Graph Laplacian \mathbf{L} associated with \mathcal{S} :

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

Learning from labeled and unlabeled data

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

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

Graph Laplacian \mathbf{L} associated with \mathcal{S} :

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

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

Learning from labeled and unlabeled data

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

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

Graph Laplacian \mathbf{L} associated with \mathcal{S} :

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

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

Decision boundary $f(\mathbf{x})$: $\mathbf{f} = [f(\mathbf{x}_i), \quad i = 1, \dots, m]^T$

Learning from labeled and unlabeled data (cont.)

Regularization framework for function learning:

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

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

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

Learning from labeled and unlabeled data (cont.)

Regularization framework for function learning:

$$f^* = \operatorname{argmin}_{f \in \mathcal{H}_k} \sum_{i=1}^I V(x_i, y_i, f) + \gamma_A \|f\|_A^2 + \gamma_I \|f\|_I^2$$

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 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.



Manifolds

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



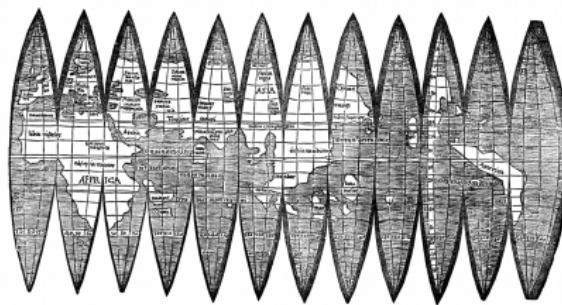
Manifolds

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



Manifolds

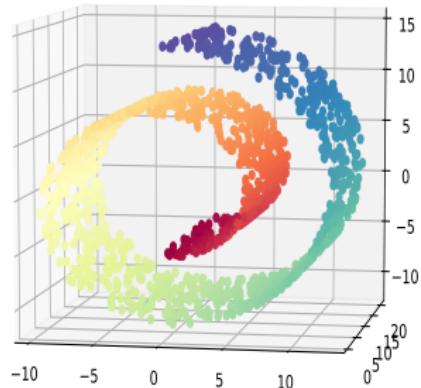
A manifold is a topological space that on a small enough scale resembles the Euclidean space.



Martin Waldseemüller, Public domain, via Wikimedia Commons

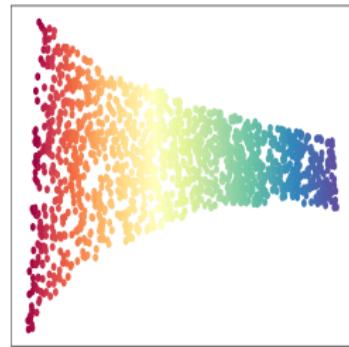
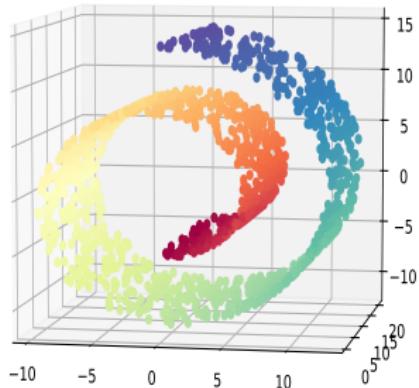
Manifold Learning

The Swiss Roll Problem



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 $\|\cdot\|_1$:

$$\|f\|_1^2 = \sum_{i=1}^m \sum_{j=i}^m w_{ij} (f(\mathbf{x}_i) - f(\mathbf{x}_j))^2 = \mathbf{f}^T \mathbf{L} \mathbf{f}$$

Learning from labeled and unlabeled data (cont.)

Representer Theorem (Kimeldorf and Wahba, 1970):

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

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

Laplacian Support Vector Machines

Constrained primal optimization problem based on the dual form:

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

subject to

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

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

Laplacian Support Vector Machines (cont.)

Lagrange function L :

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

Laplacian Support Vector Machines (cont.)

KKT condition: the gradient w. r. t. the primal variables β, β_0, ξ has to vanish

Laplacian Support Vector Machines (cont.)

KKT condition: the gradient w. r. t. the primal variables β, β_0, ξ has to vanish

- Partial derivative w. r. t. β_0 :

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

Laplacian Support Vector Machines (cont.)

KKT condition: the gradient w. r. t. the primal variables β, β_0, ξ 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 :

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

Laplacian Support Vector Machines (cont.)

Simplifying the Langrangian using the two identities above:

$$\begin{aligned}
 L(\beta, \lambda) &= \frac{1}{2} \beta^T (2\gamma_A \mathbf{K} + 2\gamma_L \mathbf{KLK}) \beta - \\
 &\quad - \sum_{i=1}^I \lambda_i \left(y_i \left(\sum_{j=1}^m \beta_j k(\mathbf{x}_i, \mathbf{x}_j) + \beta_0 \right) - 1 \right) \\
 &= \frac{1}{2} \beta^T (2\gamma_A \mathbf{K} + 2\gamma_L \mathbf{KLK}) \beta - \beta^T \mathbf{KJ}_{\mathcal{L}}^T \mathbf{Y} \lambda + \sum_{i=1}^I \lambda_i
 \end{aligned}$$

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

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

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

Laplacian Support Vector Machines (cont.)

Partial derivative w.r.t. β :

$$\frac{\partial L}{\beta} = 0 \Rightarrow (2\gamma_A K + 2\gamma K L K) \beta - K J_{\mathcal{L}}^T Y \lambda = 0$$
$$\Rightarrow \beta = (2\gamma_A I + 2\gamma K L)^{-1} J_{\mathcal{L}}^T Y \lambda$$

Laplacian Support Vector Machines (cont.)

Partial derivative w.r.t. β :

$$\begin{aligned}\frac{\partial L}{\beta} &= 0 \quad \Rightarrow \quad (2\gamma_A \mathbf{K} + 2\gamma \mathbf{KLK})\beta - \mathbf{KJ}_{\mathcal{L}}^T \mathbf{Y}\lambda = 0 \\ &\Rightarrow \quad \beta = (2\gamma_A \mathbf{I} + 2\gamma \mathbf{KL})^{-1} \mathbf{J}_{\mathcal{L}}^T \mathbf{Y}\lambda\end{aligned}$$

Note: direct relationship between parameters β and Lagrange multipliers λ

Laplacian Support Vector Machines (cont.)

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

$$\max_{\lambda \in \mathbb{R}^l} \sum_{i=1}^l \lambda_i - \frac{1}{2} \lambda^T Q \lambda$$

subject to $0 \leq \lambda_i \leq 1, \quad i = 1, \dots, l$

$$\sum_{i=1}^l \lambda_i y_i = 0$$

where

$$Q = Y J_{\mathcal{L}} K (2\gamma_A I + 2\gamma_I K L)^{-1} J_{\mathcal{L}}^T Y$$

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