

# MATHEMATICAL ASPECTS OF MACHINE LEARNING

REPORT

---

## Digit Recognition with Support Vector Machine

---

*Authors:*

Lisa GAEDKE-MERZHÄUSER

Paul KORSMEIER

Lisa MATTRISCH

Vanessa SCHRECK

# Problem Statement

Our main goal is to correctly identify handwritten digits based on the MNIST ("Modified National Institute of Standards and Technology") data set. This data set consists of 42000 gray-scale images. Each image is 28 pixels in height and 28 pixels in width. Each pixel has a single pixel-value associated with it, indicating the lightness or darkness of that pixel (kag).

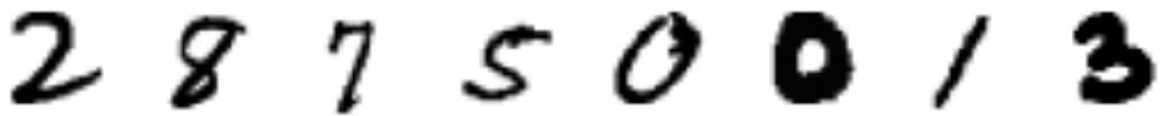


Figure 1: Visualization of eight of these images

# Support Vector Machine

Given data points  $\{x_i\}_{i=1}^l \subset \mathbb{R}^d$  with labels  $\{y_i\}_{i=1}^l \subset \{\pm 1\}$  we want to separate the different labels using a decision function of the form

$$f(x) = \text{sign}(w^T \Phi(x) + b)$$

where  $\Phi$  is the feature map as introduced in the lecture. The Support Vector Machine (SVM) is a binary classifier that aims to find the maximum margin hyperplane in the feature space, i.e. the goal is to maximize the margin while softly penalizing points that lie on the wrong side of the boundary or inside the margin (Bis06). After the dualization of this optimization problem we obtain the following equivalent quadratic program:

$$\underset{0 \leq \alpha \leq C}{\text{maximize}} \quad \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i,j=1}^l \alpha_i \alpha_j y_i y_j k(x_i, x_j)$$

with the dual variable  $\alpha$ , kernel function  $k$  and penalty  $C$ . After finding the optimal solution  $\alpha^*$  of this program we can formulate the decision function as

$$f(x) = \text{sign} \left( \sum_{i=1}^l \alpha_i^* y_i k(x_i, x_j) + b \right).$$

Note that only data points  $x_i$  with  $\alpha_i \neq 0$  appear in the decision function. Those points are called *support vectors*. By analysing primal, dual and the corresponding KKT conditions we find that any  $x_i$  with  $\alpha_i = C$  lies exactly on the margin and any  $x_i$  with  $0 < \alpha_i < C$  lies either inside the margin or on the wrong side of the margin. Additionally any data point  $x_i$  on the margin must satisfy  $y_i * f(x_i) = 1$ . We can use this property to find the parameter  $b$  (Bis06).

# Solving the optimization problem with SMO

ecoc

# Bibliography

[Bis06] BISHOP, Christopher M.: *Pattern Recognition and Machine Learning* -. 1st ed. 2006. Corr. 2nd printing 2011. Berlin, Heidelberg : Springer, 2006. – ISBN 978-0-387-31073-2

[kag] <https://www.kaggle.com/c/digit-recognizer/data>