

2.2 Linear Support Vector Machines

Machine Learning 1: Foundations

Marius Kloft (TUK)

28 Apr – 5 May 2020

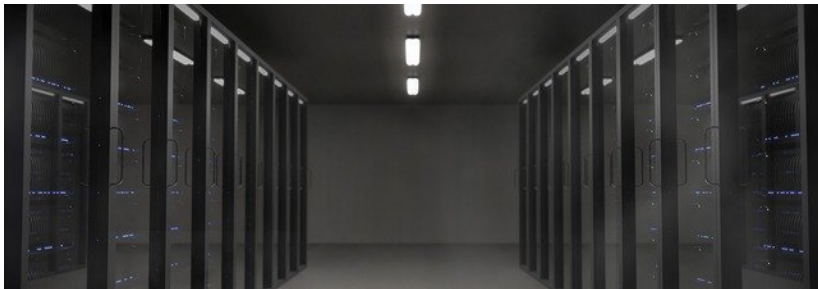
1 Teaser

2 Hard-Margin Linear SVMs

3 Soft-Margin Linear SVMs

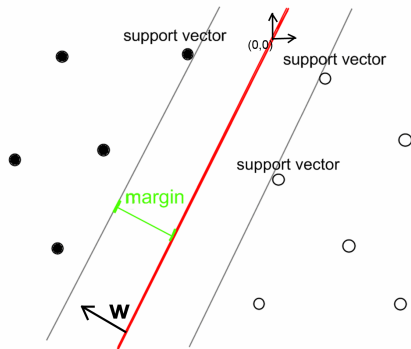
Properties of Linear SVMs: 1. Fast

- ▶ Can be trained in $O(n \cdot d)$
- ▶ Can be trained in a distributed manner (map-reduce)



Properties of Linear SVMs: 2. Simple

Geometrical interpretation:

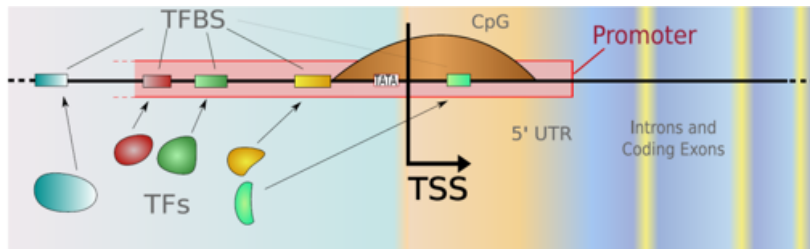


Properties of Linear SVMs: 3. **Accurate** (in ca. 50% of the data out there)

State of the art in many application areas, e.g.:

Gene Finding

Find in the DNA the positions that impact virtually all important inherited properties of you!
(e.g., intelligence, height, visual appearance, etc.)



SVM State of the Art in...

Ad Click Prediction

Predict the ad that has the highest probability of being clicked by the user

The screenshot shows a web browser window with the address bar displaying 'mloftmaths.uni-kl.de'. The website layout includes a sidebar on the left with a 'Trending' section containing three items: 'Could UV Light be the cure to COVID-19?', 'Click HERE, to learn Machine Learning without any Maths', and 'Does your mask suit your eye color?'. The main content area is titled 'Prof. Dr. Marius Kloft' and features a profile picture, contact information for Technische Universität Kaiserslautern, and a list of research interests including statistical machine learning and applications in genomics and chemical engineering. A 'News' section at the bottom mentions a job chair for HAN 20 and H5DATS 20 and ECOM 20.

Trending

Could UV Light be the cure to COVID-19?

Click HERE, to learn Machine Learning without any Maths

Does your mask suit your eye color?

Prof. Dr. Marius Kloft

Technische Universität Kaiserslautern
Department of Computer Science
Building 39, Room 214
P.O. Box 3040
67033 Kaiserslautern

Contact
m.kloft@informatik.uni-kl.de (job applications)
m.kloft@informatik.uni-kl.de (secretary, appointments and administration)
m.kloft@informatik.uni-kl.de (academic matters)

Secretary
Tobias Haas

[Bio | Research | Groups | Jobs | Teaching | Publications | Activities]

Short Bio

Since 2017 Marius Kloft has been a professor of computer science at TU Kaiserslautern, Germany. Previously, he was an adjunct associate professor at the University of California, La Jolla (1919-1921), an [Lecturer \(Associate Professor\)](#) (2014), an [assistant professor](#) at MIT (2014-2017) and a [joint professor](#) at MIT (2017-2019) at the Center for Mathematical Science and Physical Sciences, and a [Senior Researcher](#) at Google, New York, working with [Jeffrey Dean](#), [Christos Demetrescu](#), and [Guy E. Blelloch](#). From 2007-2011, he was a PhD student in the machine learning program of TU Berlin, headed by [Sven Dietterich](#). He was co-advised by [Gilles Blelloch](#) and [Daniel A. Roth](#), where heaving them group at MIT. He held from 2009 to 2010, he received a master in mathematics from the University of Würzburg with a thesis in algebraic geometry.

Research Interests

Marius Kloft is interested in theory and algorithms of statistical machine learning and its applications, especially in statistical genomics and chemical engineering. He has been working on, e.g., multiple kernel learning, transfer learning, accurate detection, extreme classification, and universal learning for computer security. He co-organized workshops on these topics at [NIPS 2010](#), [2012](#), [2014](#), [2015](#), [ECML 2016](#), [ECML 2018](#), and [ICML 2019](#). His quantitative papers on machine learning were recognized by TU Berlin for the National Dissertation Award of the German Chapter of the ACM (2011). In 2019, he received the Google Mind Influential Paper 2019 Award.

News

- New chair for HAN 20 and H5DATS 20 and ECOM 20.

SVM has Generated Zillions and Zillions of Money ...

\$75 billion Revenue

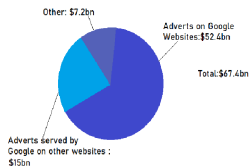
\$23 billion operating profit

The money comes from:

USA: \$34.8bn

UK: \$7.1bn

Rest: \$33.1bn

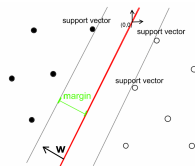


Google in 2015. Source: <http://www.bbc.co.uk/guides/z9x6bk7>

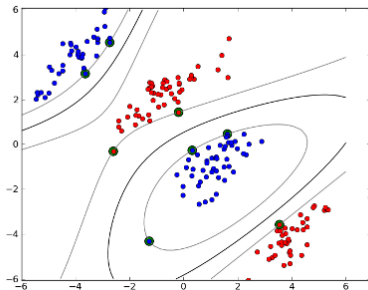
Types of (Linear) SVMs

There are two different sorts of linear SVMs:

- 1 **Hard-margin** linear SVMs
- 2 **Soft-margin** linear SVMs



Later in the course, we will also learn about **non-linear** SVMs.

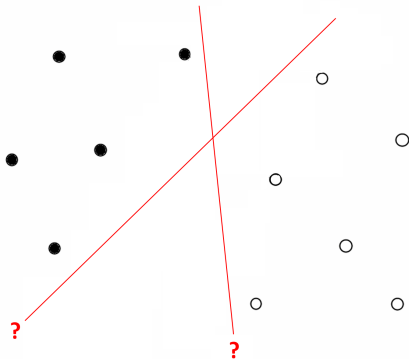


- 1 Teaser
- 2 Hard-Margin Linear SVMs
- 3 Soft-Margin Linear SVMs

Linear Support Vector Machines

Core idea:

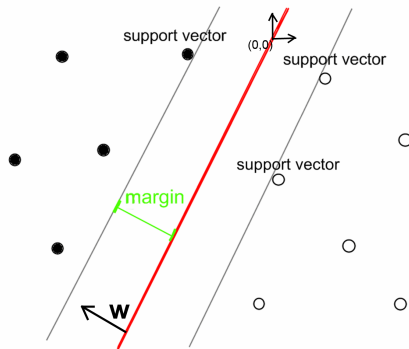
- Which hyperplane to take?



Linear Support Vector Machines

Core idea:

- ▶ Which hyperplane to take?
- ▶ The one that separates the data with the **largest margin**!



Mathematical Formalization

Maximize margin such that all data points lie outside of the margin — how can we **mathematically** describe this idea?

- ▶ Denote margin by γ
- ▶ Find hyperplane parameters \mathbf{w} and b that maximize γ
- ▶ but make sure that all positive data points lie on one side
 - ▶ a point \mathbf{x}_i with $y_i = +1$ lies on correct side of margin if:
$$d(\mathbf{x}_i, H) \geq +\gamma$$
- ▶ and all negative points on the other
 - ▶ a point \mathbf{x}_i with $y_i = -1$ lies on correct side of margin if:
$$d(\mathbf{x}_i, H) \leq -\gamma$$
- ▶ Hence, we require for all points \mathbf{x}_i ,
$$y_i \cdot d(\mathbf{x}_i, H) \geq \gamma$$

Find hyperplane H with maximal margin γ such that (s.t.) for all training points \mathbf{x}_i it holds: $y_i \cdot d(\mathbf{x}_i, H) \geq \gamma$.

Mathematical Formalization (2)

Recall:

- ▶ The hyperplane H is parameterized by \mathbf{w} and b through:

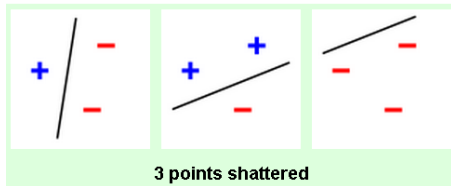
$$H = \{\mathbf{x} : \mathbf{w}^\top \mathbf{x} + b = 0\}$$

- ▶ Previous proposition: $d(\mathbf{x}, H) = \frac{1}{\|\mathbf{w}\|} (\mathbf{w}^\top \mathbf{x} + b)$

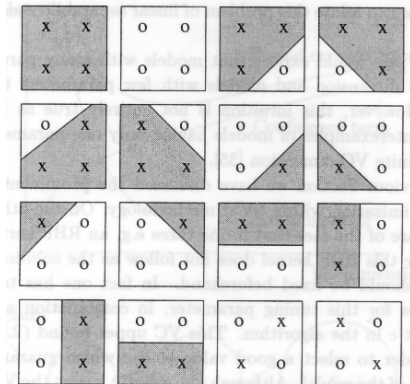
Preliminary formulation of SVM

$$\begin{aligned} & \max_{\gamma, b \in \mathbb{R}, \mathbf{w} \in \mathbb{R}^d} \quad \gamma \\ & \text{s.t.} \quad y_i (\mathbf{w}^\top \mathbf{x}_i + b) \geq \|\mathbf{w}\| \gamma, \quad \forall i = 1, \dots, n \end{aligned}$$

Limitations of Hard-Margin SVMs



Any three points in the plane \mathbb{R}^2 (not lying on a line) can be “shattered” (separated) by a line (= linear classifier).

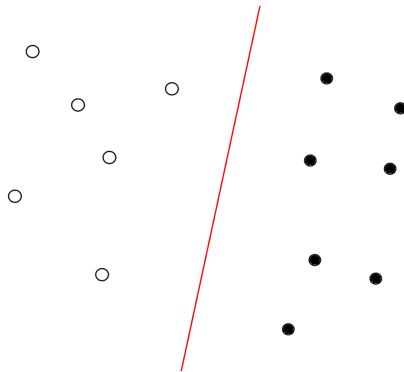


But there are configurations of four points which no hyperplane can shatter. More generally:

Any $n + 1$ points in \mathbb{R}^n (not lying in a hyperplane) can be “shattered” by a hyperplane. But there are configurations of $n + 2$ points which no hyperplane can shatter.

Limitations Hard-Margin Linear SVMs (continued)

Another Problem is that of outliers potentially corrupting the SVM:



- 1 Teaser
- 2 Hard-Margin Linear SVMs
- 3 Soft-Margin Linear SVMs

Remedy: **Soft**-Margin Linear SVMs

Core idea:

- ▶ Introduce for each input \mathbf{x}_i a **slack variable** $\xi_i \geq 0$ that allows for some (slight violations of the margin separation):

Linear Soft-Margin SVM

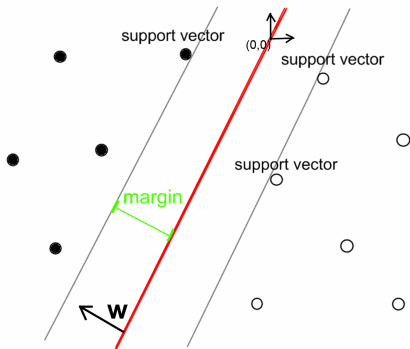
$$\begin{aligned} \max_{\gamma, b \in \mathbb{R}, \mathbf{w} \in \mathbb{R}^d, \xi_1, \dots, \xi_n \geq 0} \quad & \gamma - C \sum_{i=1}^n \xi_i \\ \text{s.t.} \quad & y_i \left(\mathbf{w}^\top \mathbf{x}_i + b \right) \geq \|\mathbf{w}\| \gamma - \xi_i, \quad \forall i = 1, \dots, n \end{aligned}$$

- ▶ minimizes $\sum_{i=1}^n \xi_i$, to allow for “slight” violations of the margin
- ▶ $\sum_{i=1}^n \xi_i$ is the total amount of violations of the training points lying inside the margin (measured in distances to the margin)
- ▶ $C > 0$ is a trade-off parameter (to be set in advance): the higher C , the more we penalize violations

Why the Name 'Support Vector Machine'?

Denote by γ^* and \mathbf{w}^* the optimal arguments from previous slide

Def.: All vectors \mathbf{x}_i with $y_i \cdot d(\mathbf{x}_i, H(\mathbf{w}^*, b^*)) \leq \gamma^*$ (i.e., lying inside the tube) are called **support vectors**.



The SVM depends only on the support vectors: all other points can be removed from the training data (no impact on classifier)

PanOpto Quiz

Alternatively, one could consider a variation of the SVM, where the penalty term $C \sum_{i=1}^n \xi_i$ is replaced by $C \sum_{i=1}^n \xi_i^2$. Would this be a reasonable SVM? (i.e., could this work similarly well as the original soft-margin SVM?)

Let us again remove the restriction $\xi_i \geq 0 \forall i$ in the SVM's maximization. Let us remove from our training set all data points that lie strictly outside the margin. Will the decision boundary change?

SVM training

How can we train SVMs, that is, how to solve the minimization task?

Next Week: Convex Optimization Problems

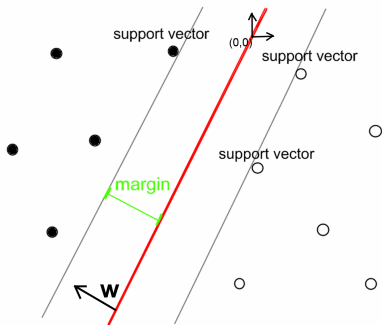
It is known from decades of research in numerical mathematics that so-called **convex optimization problems** (to be introduced in detail next week) can be solved very efficiently.

Will show: we can view the SVM as a convex optimization problem.

Conclusion

Linear Support Vector Machines (SVMs)

- motivated geometrically



Mathematical formalization of this picture:

$$\begin{aligned} \max_{\gamma, b \in \mathbb{R}, \mathbf{w} \in \mathbb{R}^d} \quad & \gamma \\ \text{s.t.} \quad & y_i \left(\mathbf{w}^\top \mathbf{x}_i + b \right) \geq \|\mathbf{w}\| \gamma, \quad \forall i = 1, \dots, n \end{aligned}$$

Suggested Reading

[1]: The Elements of Statistical Learning, Sections 4.5 and
Section 12.2

Refs I



T. Hastie, R. Tibshirani, and J. Friedman, The elements of statistical learning, 2nd edition. Springer series, 2009.