

**Support Vector Machines s kernelmi**

SEMESTRÁLNA PRÁCA

Vypracoval(a): **Adam Hajro**

Študijná skupina: **5ZYI25**

Predmet: **Dáta, informácie, znalosti**

Cvičiaci: *Ing. Lukáš Falát, PhD.*

Obsah

[1. Support vector machines 3](#_Toc125383215)

[1.1. Čo sú vlastne zač? 3](#_Toc125383216)

[1.2. Lineárne dáta 3](#_Toc125383217)

[2. Požitý dataset 5](#_Toc125383218)

[3. Postup implementácie vo vybranom jazyku 5](#_Toc125383219)

[4. Výsledky 5](#_Toc125383220)

# Support vector machines

## Čo sú vlastne zač?

“In machine learning, support vector machines (SVMs, also support vector networks[1]) are supervised learning models with associated learning algorithms that analyze data for classification and regression analysis. Developed at AT&T Bell Laboratories by Vladimir Vapnik with colleagues (Boser et al., 1992, Guyon et al., 1993, Cortes and Vapnik, 1995,[1] Vapnik et al., 1997[citation needed]) SVMs are one of the most robust prediction methods, being based on statistical learning frameworks or VC theory proposed by Vapnik (1982, 1995) and Chervonenkis (1974). Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier (although methods such as Platt scaling exist to use SVM in a probabilistic classification setting). SVM maps training examples to points in space so as to maximise the width of the gap between the two categories. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall.

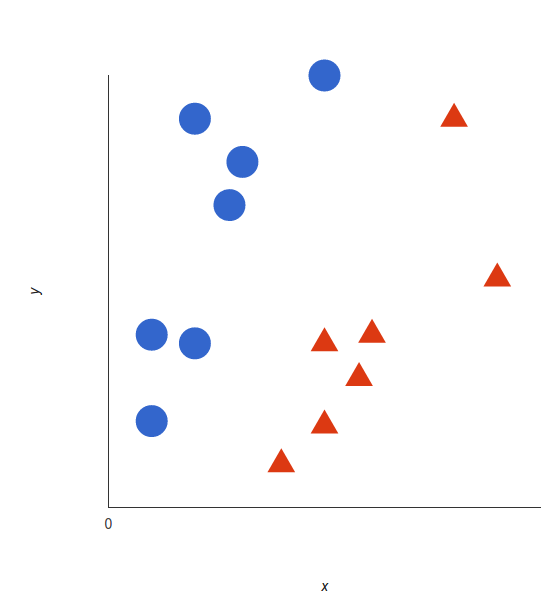
Jednoduché vysvetlenie:

„*In machine learning, support vector machines are supervised learning models with associated learning algorithms that analyse data used for classification and regression analysis. Mostly are being used in classification problems*”

## Lineárne dáta

Základy fungovania SVM je možné pochopiť týmto jednoduchým príkladom.

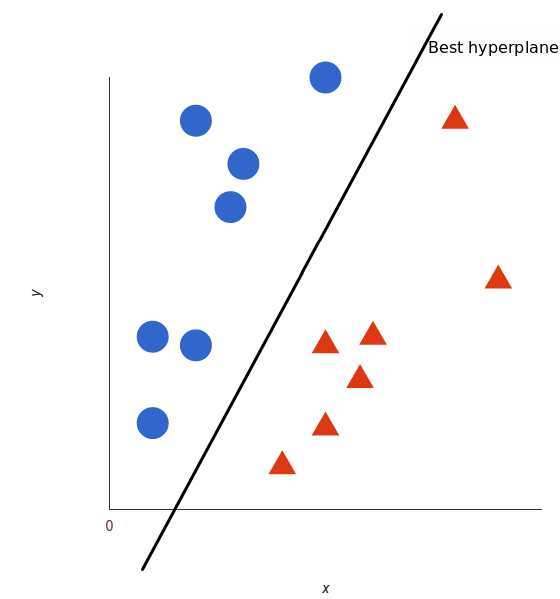
Predstavme si že máme

* dva tagy: **červený** a **modrý**
* a definované dáta sú na dvoch osách: **x** a **y**

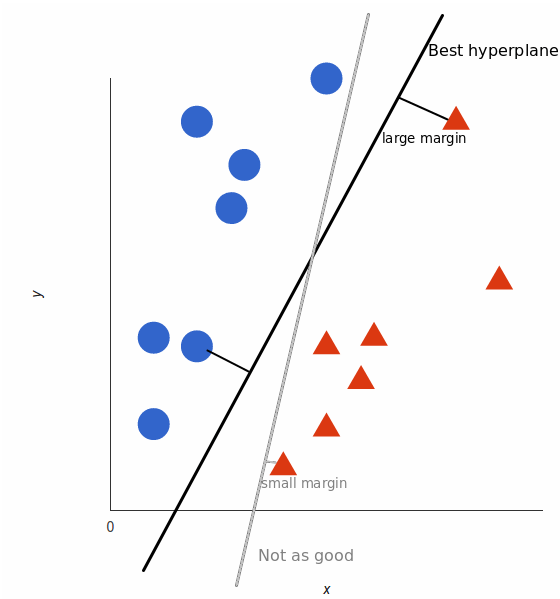
Chceme nájsť **jednoznačný** classifier, kt. na základe **x**, **y** súradníc určí či je výsledok **červený** alebo **modrý.**

SVM zoberie dátové body a vyznačí tzv. **hyperplane** (v 2D svete je to čiara), kt. najlepšie rozdeľuje tagy.

Táto čiara je rozhodujúca hranica:

* všetko, čo padne na jej **JEDNU** stranu sa klasifikuje ako **MODRÉ**
* všetko, čo padne na jej **DRUHÚ** stranu sa klasifikuje ako **ČERVENÉ**

Potom tu nastáva otázka že kt. **hyperplane**/čiara je najlepšia?

* pre SVM je to tá, kt. **maximalizuje** *margins* z obidvoch tagov
* resp. hyperplane kt. vzdialenosť k najbližšiemu prvku každého tagu je najväčšia

# Požitý dataset

Kategorický dataset ktorý určuje že či si daný používateľ zakúpil konkrétny produkt.

Viacej info na: <https://www.kaggle.com/datasets/rakeshrau/social-network-ads>

# Postup implementácie vo vybranom jazyku

Celý postup sa nachádza v zdrojovom kóde v súbore **support-vector-machines.R**

# Výsledky

