Support Vector Machine (SVM) classifier

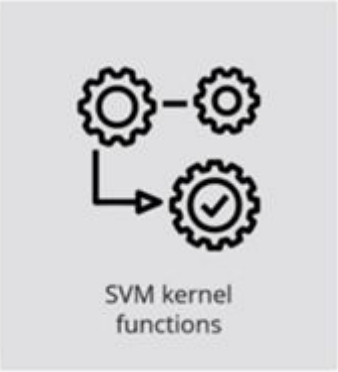
Volha Shyrayeva, FAMCS,
08.10.2019

Contents

1. How SVM works : theory and visual explanation.
2. Introduction to Non-Linear SVM tricks.
3. Tuning parameters of SVM.

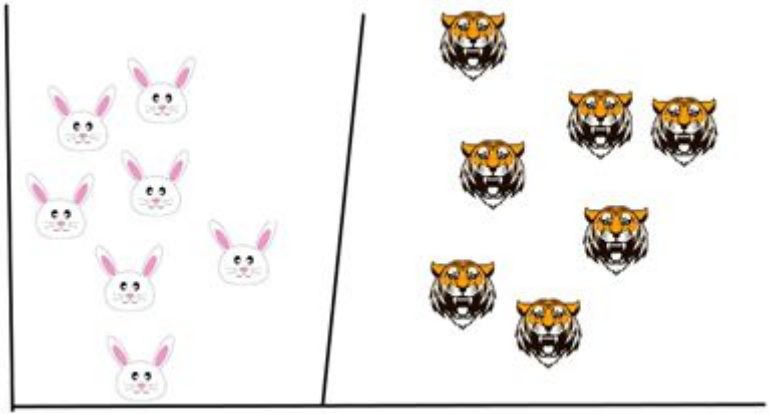
1. How SVM works : theory

SVM is a supervised classification method that separates data using hyperplanes.

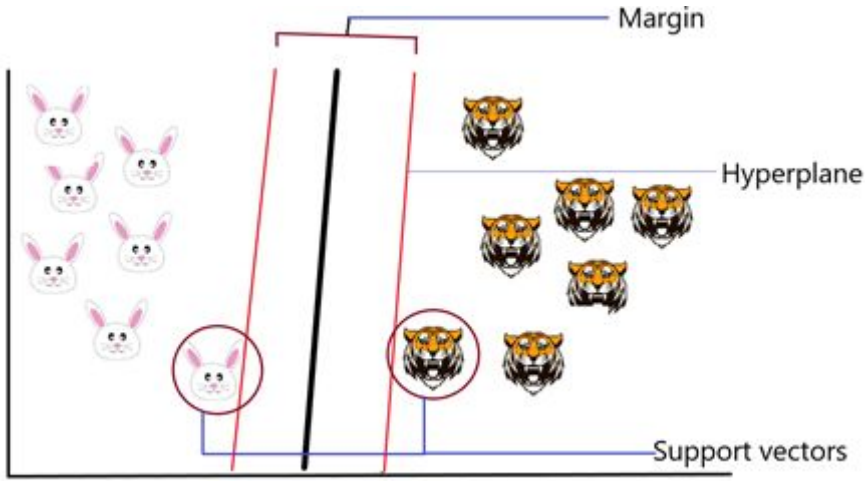


https://medium.com

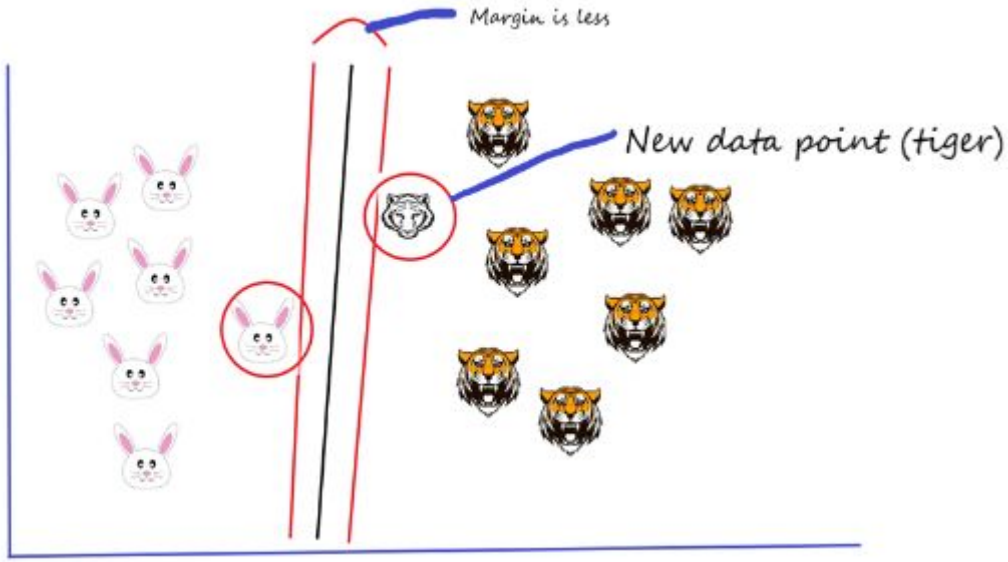
https://medium.com



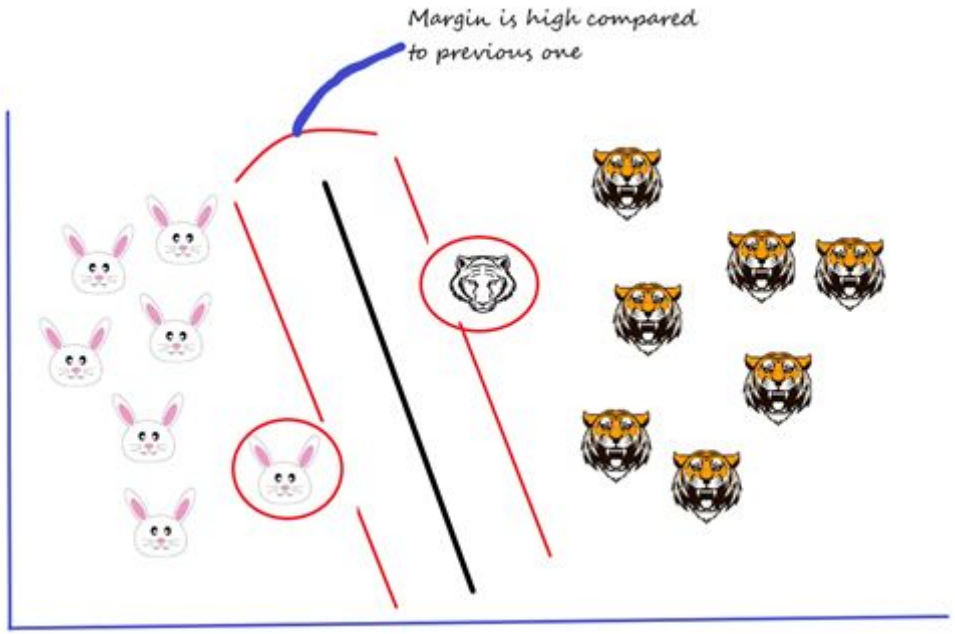
https://medium.com



https://medium.com



https://medium.com



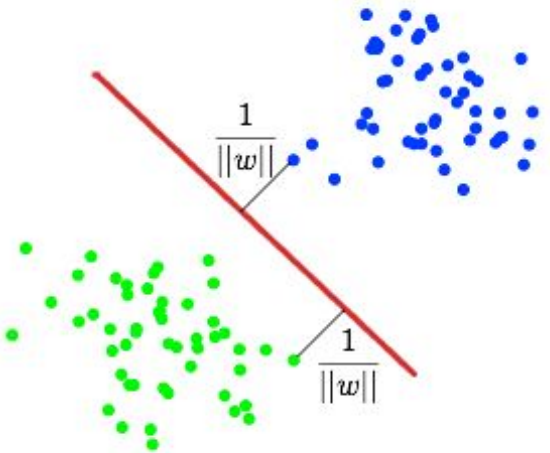
https://medium.com

Finding the most suitable hyperplane

$$(x_1, y_1), \ldots, (x_m, y_m), \; x_i \in \mathbb{R}^n, \; y_i \in \{-1, 1\}.$$

$$F(\mathbf{x}) = \operatorname{sign}\left(\langle \mathbf{w}, \mathbf{x} \rangle + b\right)$$

$$\langle \mathbf{w}, \mathbf{x} \rangle + b = 0$$



https://habr.com/ru/post/105220/

$$\begin{cases} \arg \min\limits_{\mathbf{w},b} ||w||^2, \\ y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geqslant 1, \quad i = 1, \ldots, m. \end{cases}$$

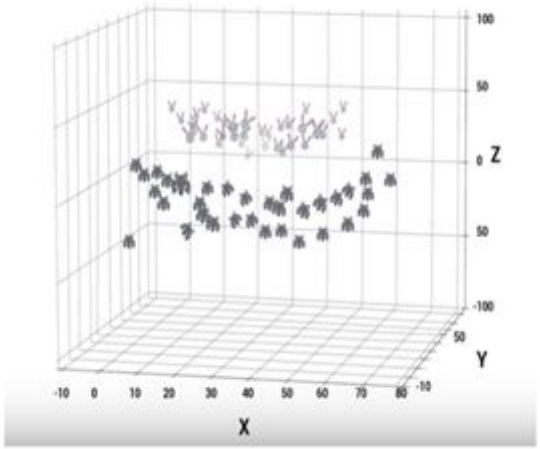
$$\begin{cases} \mathbf{L}(\mathbf{w}, \mathbf{b}; \lambda) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{n} \lambda_\mathbf{i}(c_i((\mathbf{w} \cdot \mathbf{x_i}) - b) - 1) \to \min_{w,b} \max_{\lambda} \\ \lambda_\mathbf{i} \geq 0, \quad 1 \leq i \leq n \end{cases}$$

https://ru.wikipedia.org/wiki/Метод_опорных_векторов

https://medium.com

Introduction to non-linear SVMs

https://medium.com



https://medium.com

$$\varphi : \mathbb{R}^n \to X \qquad F(\mathbf{x}) = \operatorname{sign}\left(\langle \mathbf{w}, \varphi(\mathbf{x}) \rangle + b\right)$$

$$\mathbf{k}(\mathbf{x}, \mathbf{x}') = \langle\, \varphi(\mathbf{x}), \varphi(\mathbf{x}') \,\rangle$$

$$\mathbf{k}(\mathbf{x}, \mathbf{x}') = (\langle \mathbf{x}, \mathbf{x}' \rangle + \text{const})^d$$

$$\mathbf{k}(\mathbf{x}, \mathbf{x}') = e^{-\gamma ||\mathbf{x} - \mathbf{x}'||^2}, \gamma > 0.$$

$$\mathbf{k}(\mathbf{x}, \mathbf{x}') = e^{-\frac{||\mathbf{x} - \mathbf{x}'||^2}{2\sigma^2}}.$$

$$\mathbf{k}(\mathbf{x}, \mathbf{x}') = \tanh(\kappa \langle \mathbf{x}, \mathbf{x}' \rangle + c), \ \kappa > 0, c < 0.$$

https://habr.com/ru/post/105220/

Tuning parameters of SVM

$$\begin{cases} \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i \to \min_{w,b,\xi_i} \\ c_i(\mathbf{w}\cdot\mathbf{x_i} - b) \geq 1 - \xi_i, \quad 1 \leq i \leq n \\ \xi_i \geq 0, \quad 1 \leq i \leq n \end{cases}$$

Cupcakes

versus

Muffins

"a cupcake is just a muffin with frosting"

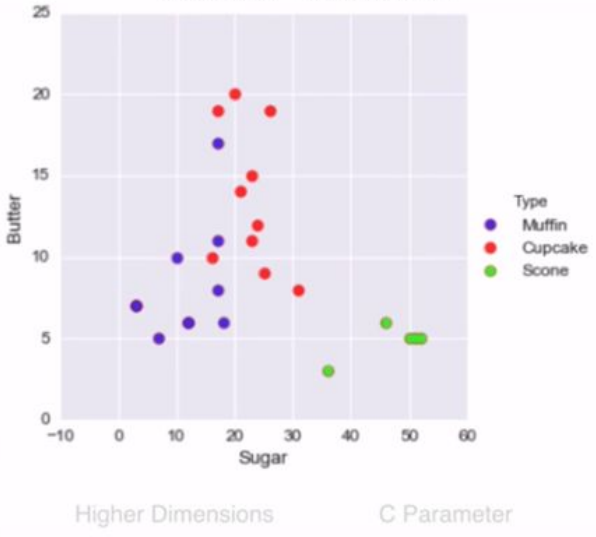
"a muffin is just a cupcake with random bits of stuff in it"

https://medium.com

Datset

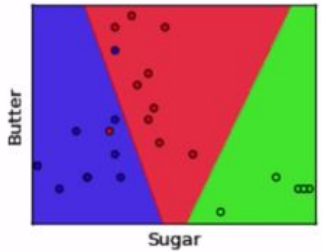
Type	Flour	Milk	Sugar	Butter	Egg	Baking Powder	Vanilla	Salt
Muffin	55	28	3	7	5	2	0	0
Muffin	47	24	12	6	9	1	0	0
Muffin	47	23	18	6	4	1	0	0
Muffin	50	25	12	6	5	2	1	0
Muffin	55	27	3	7	5	2	1	0
Muffin	54	27	7	5	5	2	0	0
Muffin	47	26	10	10	4	1	0	0
Muffin	50	17	17	8	6	1	0	0
Muffin	50	17	17	11	4	1	0	0
Cupcake	39	0	26	19	14	1	1	0
Cupcake	34	17	20	20	5	2	1	0
Cupcake	39	13	17	19	10	1	1	0
Cupcake	38	15	23	15	8	0	1	0
Cupcake	42	18	25	9	5	1	0	0
Cupcake	36	14	21	14	11	2	1	0
Cupcake	38	15	31	8	6	1	1	0
Cupcake	36	16	24	12	9	1	1	0
Cupcake	34	17	23	11	13	0	1	0

Multi Classes



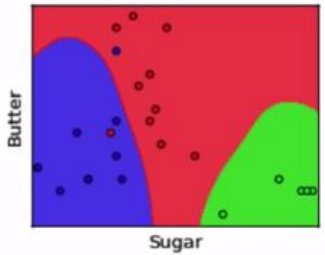
https://medium.com

Kernel Trick: Comparison



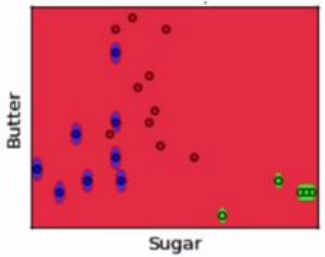
Kernel: Linear
C: 1

Muffin
Cupcake
Scone



Kernel: RBF
C: 1
Gamma: 2^-5

Small Gamma:
Less complexity



Kernel: RBF
C: 1
Gamma: 2^1

Large Gamma:
More complexity

Higher Dimensions C Parameter Multiple Classes **Kernel Trick**

https://medium.com

Thank you.