Recap

10 July 2023 14:58

→ Intro

→ Hard margin SVM

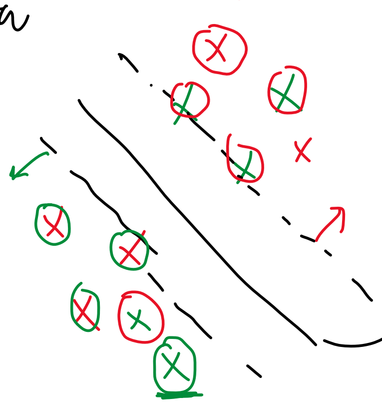
→ Problems

→ Soft margin SVM

Problems with Hard Margin SVM → Soft margin SVM

08 July 2023 08:07

→ linear data



→ big prob →

x_{1i}	x_{2i}	y_i
X_1	X_2	Y
x_{11}	x_{21}	y_1
x_{12}	x_{22}	y_2
x_{13}	x_{23}	y_3

$AX + By + C = 0$

$$\underset{A,B,C}{\operatorname{argmax}} \frac{2}{\sqrt{A^2+B^2}}$$ such that $\boxed{y_i(Ax_{1i} + Bx_{2i} + C) \geq 1}$ for all x_i

Hard margin

flexible → soften → Soft margin SVM

Slack Variable → ξ → misclassification score

08 July 2023 19:19

Hinge loss

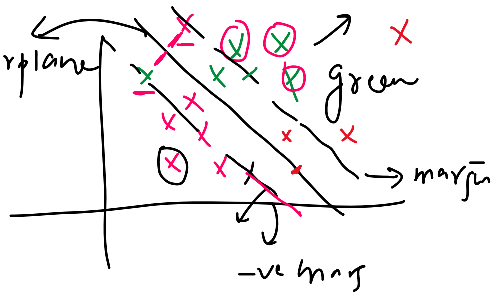
The concept of slack variables was introduced by Vladimir Vapnik in 1995 and is used in the formulation of the "soft-margin" SVM to handle cases where data is not linearly separable, or when one allows for some degree of error in classification.

Mathematically, for each data point i, a slack variable $\xi i \geq 0$ is introduced. The slack variable ξi measures the degree of misclassification of the data point xi.

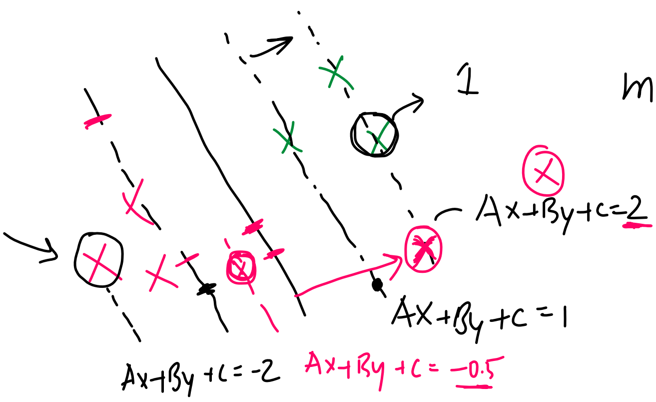
- $\xi i = 0$ if xi is on the correct side of the margin.
- $0 < \xi i < 1$ if xi is on the correct side of the hyperplane but on the wrong side of the margin.
- $\xi i \geq 1$ if xi is on the wrong side of the hyperplane, i.e., it is misclassified.

100 rows

$\xi_i \rightarrow \boxed{\xi_i \geq 0}$



Hinge loss (slack) $= \boxed{\max(0, 1 - y_i(Ax_{1i} + Bx_{2i} + C))}$ ②



1

$\max(0, 1 - 1(2))$

$\max(0, -1) = 0$

0.5

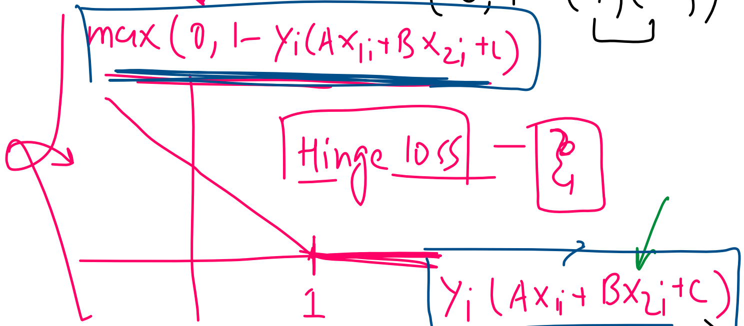
$\max(0, 1 - (-1)(-0.5))$

$\max(0, 0.5) = 0.5$

$\max(0, 1 - (-1)(-2)) = 0$

$\max(0, 1 - (-1)(2))$

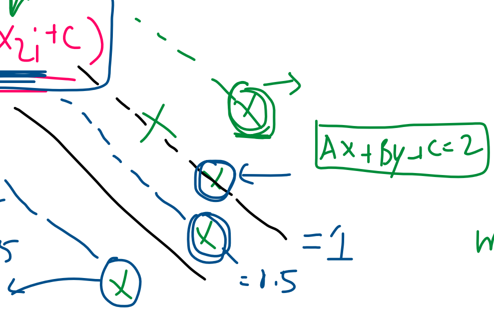
$\max(0, 3) = 3$



$1 \times 0.5 = 0.5$

$1 \times 1 = 1$

-0.5



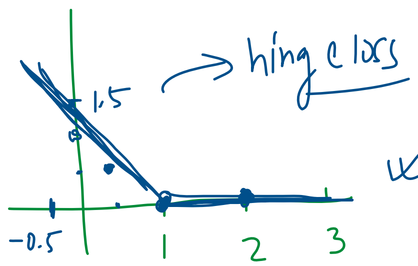
$(1 \times 2) = 2$

$(2, 0)$

$\max(0, 1-2) \to 0$

$(0.5, 0.5)$ $(1, 0)$

$(-0.5, 1.5)$



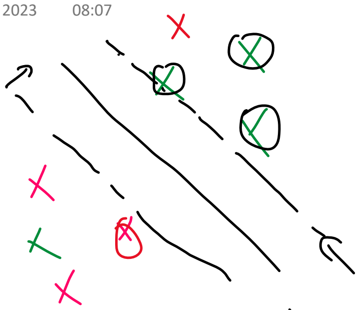
$\max(0, 1-1) \Rightarrow \max(0,0) = 0$

$\max(0, 1-0.5) = 0.5$

$\max(0, 1-(-0.5)) = 1.5$

Soft Margin SVM

08 July 2023 08:07



$x \to \frac{1}{x}$

$$\underset{A,B,C}{\arg\max} \boxed{\frac{2}{\sqrt{A^2+B^2}}}$$

such that $\boxed{y_i(Ax_{1i}+Bx_{2i}+c) \geq 1}$

for all x_i

loss function → error → reduce

change

Hard margin SVM → $\underset{A,B,C}{\arg\min} \boxed{\frac{\sqrt{A^2+B^2}}{2}}$

such that $y_i(Ax_{1i}+Bx_{2i}+c) \geq 1$

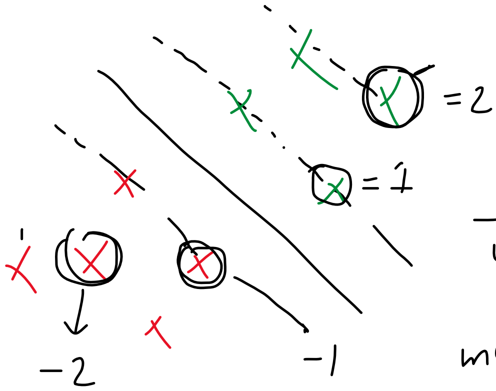
for all x_i

$$\left[\underset{A,B,C}{\arg\min} \frac{\sqrt{A^2+B^2}}{2}\right.$$

such that $\boxed{y_i(Ax_{1i}+Bx_{2i}+c) \geq 1-\xi_i}$

valid → for all x_i

such that $\boxed{\xi_i \geq 0}$



$2 \geq 1$

$-1(-2) \geq 1-\max(0, 1-2)$

$\max(0,-1) = 0$

$1(2) \geq 1-\max(0, 1-1(2))$

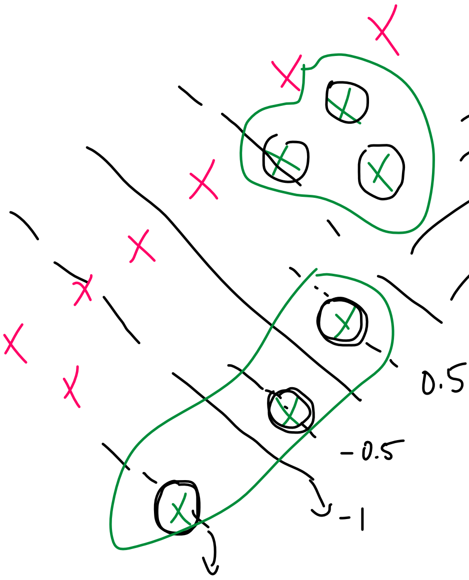
$2 \geq 1$

$-1(-1) \geq 1-\max(0,1-1)$

$1 \geq 1$

$1(1) \geq 1-\max(0,1-1)$

$1 \geq 1$



$y_i(Ax_{1i}+Bx_{2i}+C) \geq 1-\xi_i$

$1(0.5) \geq 1-\max(0, 1-0.5)$

$0.5 \geq 1-\underbrace{\max(0, 0.5)}_{0.5}$

$0.5 \geq 0.5$

true

$1(-0.5) \geq 1-\max(0, 1+0.5)$

$-0.5 \geq 1-\underbrace{\max(0, 1.5)}_{1.5}$

-2

$\underbrace{1(-2)}_{} \geq 1 - \max(0, 1+2)$

$-2 \geq 1 - 3$

$\boxed{-2 \geq -2}$ true

true

$-0.5 \geq 1 - \max(0, \underbrace{1.5}_{1.5})$

$-0.5 \geq 0.5$

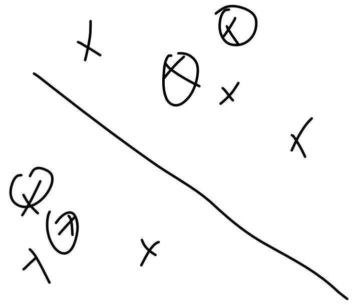
-0.5

$$\boxed{y_i(Ax_{1i} + Bx_{2i} + c) \geq 1 - \xi_i}$$

→ it is allowing all the points

All true condition

→ this is no more a constraint



Hard margin
↳ flexibility

max

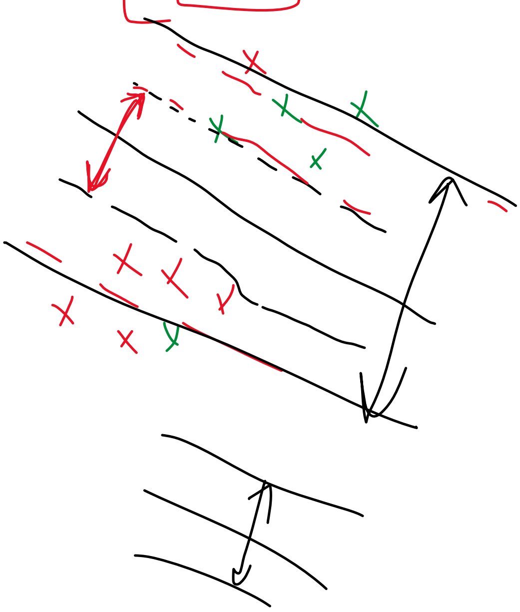
min

$$\underset{\boxed{A, B, C}}{\text{argmin}} \quad \frac{\sqrt{A^2 + B^2}}{2}$$

such that

new constraint

$$\left\{\begin{array}{l} y_i(Ax_{1i} + Bx_{2i} + c) \geq 1 - \xi_i \\ \text{for all } x_i \\ \text{such that } \xi_i \geq 0 \end{array}\right.$$



$$\underset{\boxed{A, B, C}}{\text{argmin}} \quad \boxed{\frac{\sqrt{A^2 + B^2}}{2}} + \boxed{\sum_{i=1}^{n} \xi_i}$$

maximize margin

→ reduce misclassification

such that

$$\left.\begin{array}{l} y_i(Ax_{1i} + Bx_{2i} + c) \geq 1 - \xi_i \\ \text{for all } x_i \\ \text{such that } \xi_i \geq 0 \end{array}\right\}$$

→ avg misclassi

$$\operatorname*{argmin}_{A,B,C} \left\{ \frac{\sqrt{A^2+B^2}}{2} + \frac{1}{n}\sum_{i=1}^{n} \xi_i \right\}$$

such that

$$y_i(Ax_{1i}+Bx_{2i}+C) \geq 1-\xi_i$$

for all x_i

and $\xi_i \geq 0$

Introduction of C

08 July 2023 18:44

tuning knob

hyperparameter

data

$\beta_1 x_{1i} + \beta_2 x_{2i} + \beta_0$

$\beta_1 \ \beta_2 \ \beta_0$

$$\underset{A,B,C}{\operatorname{argmin}} \ (1) \quad \frac{\sqrt{A^2+B^2}}{2} + \boxed{C}\ \frac{1}{n}\sum_{i=1}^{n} \xi_i$$

such that

$$y_i(Ax_{1i}+Bx_{2i}+C) \geq 1-\xi_i$$

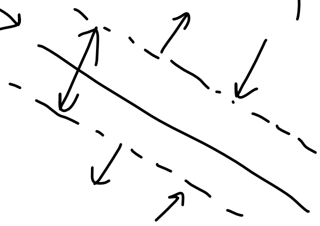
for all x_i

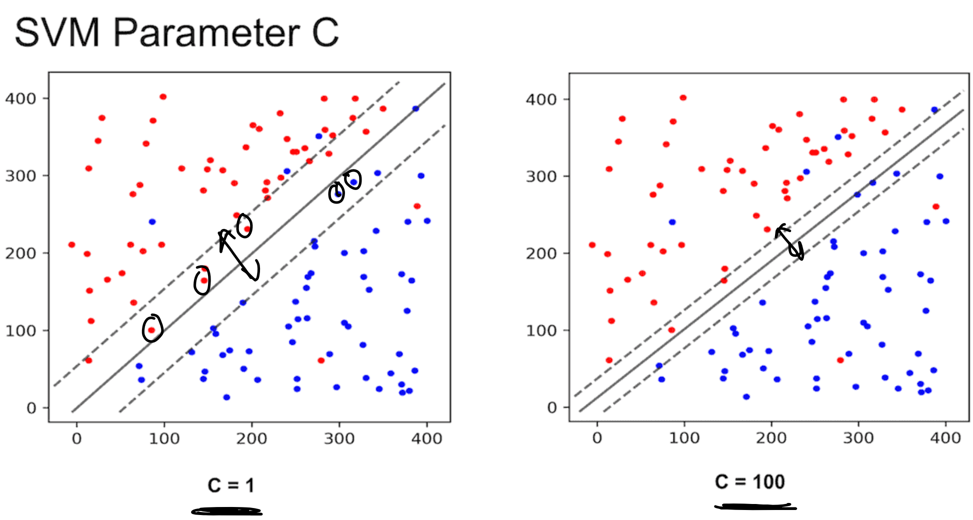
and $\xi_i \geq 0$

$\boxed{C>0}$

$\boxed{C=1000}$

$C=0.001$

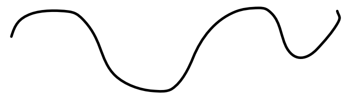




Bias Variance Tradeoff

08 July 2023 08:07

$$\underset{A\,B\,C}{\operatorname{argmin}} \; \frac{\sqrt{A^2+B^2}}{2} + C\,\frac{1}{n}\sum_{i=1}^{n} \xi_i$$



→ C high → over fitting (low bias high variance)

→ C low → underfitting (high bias low variance)

Code Example

08 July 2023 08:08

Relationship with Logistic Regression

11 July 2023 19:49

$$\underset{\beta_0 \beta_1 \beta_2}{\operatorname{argmin}} \frac{\sqrt{\beta_1^2 + \beta_2^2}}{2} + C \frac{1}{n} \sum_{i=1}^{n} \xi_i$$

loss function → hinge loss

Constrain → loss func

regu

$$\|\beta\|^2$$

$$C \propto \frac{1}{\lambda}$$

L2reg

$$\underset{\beta_0 \beta_1 \beta_2}{\operatorname{argmin}} \; \text{log loss} + \lambda \left(\sqrt{\beta_1^2 + \beta_2^2} \right)$$

regu

$$\lambda \propto \frac{1}{C}$$

$$-\frac{1}{n} \sum_{i=1}^{n} y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)$$