Soft Margin SVM

Thursday 2 May 2024 12:39 PM

→ The problem with hard margin SVM:

$$\underset{ABC}{\arg\max}\left(\frac{2}{\sqrt{A^2+B^2}}\right) \text{ s.t. } \left\{ y_i(Ax_{i1}+Bx_{i2}+C) \geq 1 \right\}$$

for all x_i

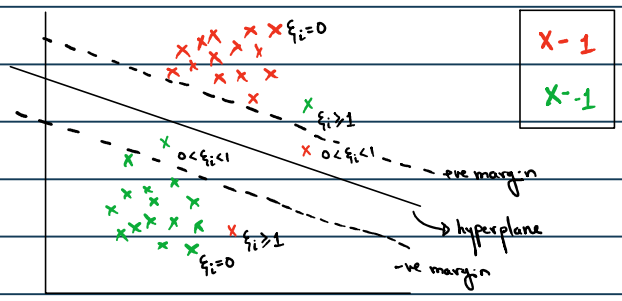
→ This constraint makes hard margin SVM inapplicable in most scenarios.

→ We need to 'soften' the constraint

→ The concept of slack variable 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.

→ for each data point, a slack variable ξ_i is introduced.

→ ξ_i measures the degree of misclassification of the data point x_i



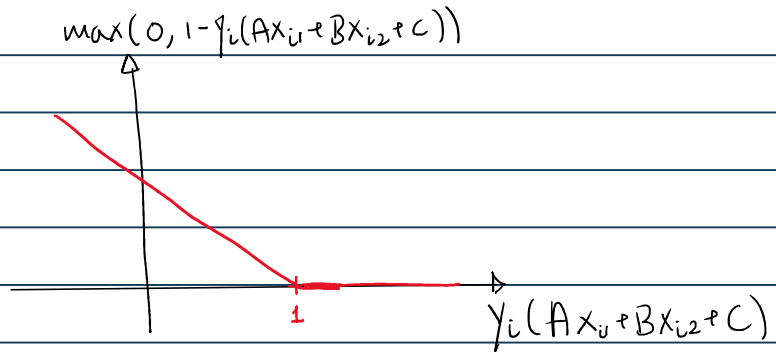
X - 1

X - -1

→ ξ_i 'misclassification score'

↳ Hinge loss

Hinge loss = $\max(0, 1-y_i(Ax_{i1}+Bx_{i2}+C))$



modified problem (overcorrection)

$$\underset{ABC}{\arg\min}\left(\frac{\sqrt{A^2+B^2}}{2}\right) \text{ s.t. } \left\{ y_i(Ax_{i1}+Bx_{i2}+C) \geq 1-\xi_i \right\}$$

for all x_i

s.t. $\xi_i \geq 0$

→ it is allowing all points / All true condition / "no more a constraint"

→ in hard margin SVM the support vectors helped in deciding the margins. But now we have modified the constraint in a way that any point can lie anywhere, there is nothing stopping the margins from increasing infinitely. There is no criteria to stopage for the margins in this modified problem.

→ addition of this term acts as constraint

addition of this term acts as constraint

$$\underset{ABC}{\arg\min}\left(\frac{\sqrt{A^2+B^2}}{2}+\frac{1}{n}\sum_{i=1}^{n}\xi_i\right) \text{ s.t. } \left\{ y_i(Ax_{i1}+Bx_{i2}+C) \geq 1-\xi_i \right\}$$

for all x_i s.t. $\xi_i \geq 0$

- $\frac{\sqrt{A^2+B^2}}{2}$ → maximize margin
- $\frac{1}{n}\sum_{i=1}^{n}\xi_i$ → reduce misclassification

opposite forces → goal is to find the balance

$C > 0$, hyperparameter, allows us to decide which "force" to prioritize more

$C \uparrow$ more focus to reduce misclassification i.e. ↓ margin vice versa.

$$\underset{ABC}{\arg\min}\left(\frac{\sqrt{A^2+B^2}}{2}+C\frac{1}{n}\sum_{i=1}^{n}\xi_i\right) \text{ s.t. } \left\{ y_i(Ax_{i1}+Bx_{i2}+C) \geq 1-\xi_i \right\}$$

for all x_i s.t. $\xi_i \geq 0$

- $\frac{\sqrt{A^2+B^2}}{2}$ → maximize margin
- $\frac{1}{n}\sum_{i=1}^{n}\xi_i$ → reduce misclassification

opposite forces → goal is to find the balance

Bias Varaiance tradeoff

↑ C → overfitting (low bias high variance)

↓ C → underfitting (high bias low variance)

Relationship with Logistic Regression

$$\underset{\beta_0\beta_1\beta_2}{\arg\min}\left(\frac{\sqrt{\beta_1^2+\beta_2^2}}{2}+C\frac{1}{n}\sum_{i=1}^{n}\xi_i\right)$$

hinge loss

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

$\sqrt{\beta_1^2+\beta_2^2} \rightarrow \|\beta\|^2$

$\lambda \propto \frac{1}{C}$ i.e. $C \propto \frac{1}{\lambda}$

→ This also creates linear decision boundaries only.