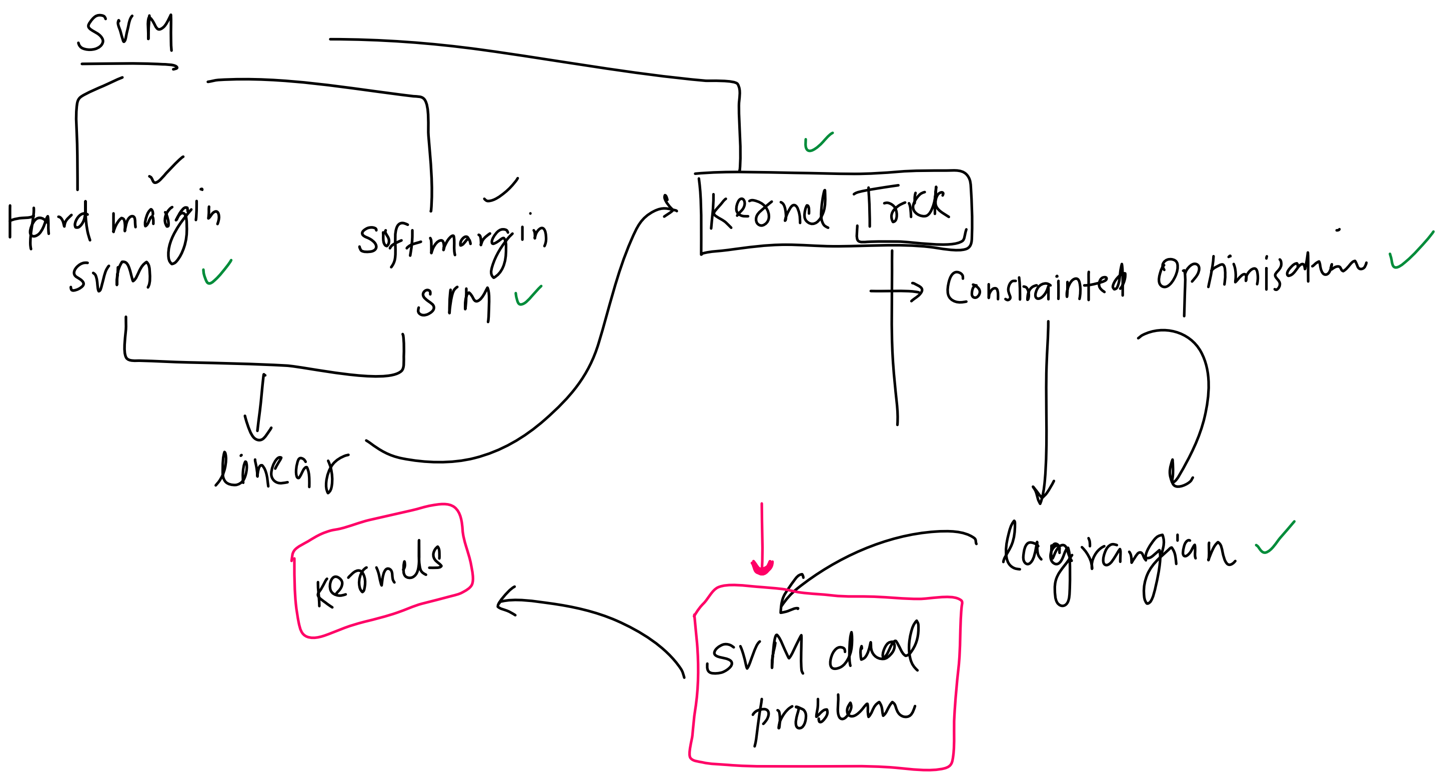
Recap

16 July 2023 00:02



SVM in n Dimensions

14 July 2023 17:42

2 dims $\rightarrow$ $\boxed{Ax + By + C}$

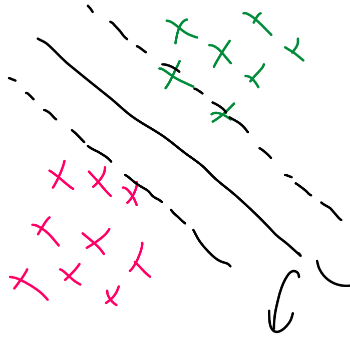
$\rightarrow \boxed{A, B, C}$

$\underline{w}$, b
↳ intercept

n-dims

$x_1 \; x_2 \; x_3 \ldots x_n \mid y$

$w_1 \; w_2 \; w_3 \ldots w_n$



$\rightarrow$ hyperplane

$\rightarrow$ Hard margin

Soft margin

$\rightarrow$ norm

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

$$w_1x_1 + w_2x_2 + w_3x_3 + \ldots + w_nx_n + b = 0$$

$$\rightarrow \boxed{w^T x + b = 0}$$

such that

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

$$\xi_i \geq 0$$

$\sqrt{w_1^2 + w_2^2 + w_3^2 + \ldots + w_n^2}$

$\rightarrow$ n-dim SVM soft margin formulation

$$\underset{w,b}{\operatorname{argmin}} \frac{\|w\|^2}{2} + C\sum_{i=1}^{n} \xi_i$$

such that

$$y_i(w^Tx + b) \geq 1 - \xi_i$$

argmin

too difficult

↳ primal $\rightarrow$ dual

$\rightarrow$ dual form $\rightarrow$ easy to solve

SVM primal $\rightleftharpoons$ SVM dual $\rightarrow$ sol^n

primal

strong duality

Constrained Optimization Problems with Inequality

16 July 2023 00:02

Constrained optimization problem equality

SVM

$$\underset{x,y}{\text{maximize}}\ f(x,y) = x^2 y$$

such that $x^2 + y^2 = 1$

Karush Kuhn Tucker Conditions (KKT conditions)

16 July 2023 11:55

They generalize the method of Lagrange multipliers to handle inequality constraints. In the context of support vector machines (SVMs) and many other optimization problems, the KKT conditions play a key role in deriving the dual problem from the primal problem.

The KKT conditions are:

1. Stationarity: The derivative of the Lagrangian with respect to the primal variables, the dual variables associated with inequality constraints, and the dual variables associated with equality constraints are all zero.
2. Primal feasibility: All the primal constraints are satisfied.
3. Dual feasibility: All the dual variables associated with inequality constraints are nonnegative.
4. Complementary slackness: The product of each dual variable and its associated inequality constraint is zero. This means that at the optimal solution, for each constraint, either the constraint is active (equality holds) and the dual variable can be nonzero, or the constraint is inactive (strict inequality holds) and the dual variable is zero.

$$\min_x \; f(x) = x^2 \quad \text{such that} \quad x - 1 \leq 0$$

inequality

SVM dual problem

primal form

dual (α)

x

$$f(x) = x^2 \quad \min_x$$

$$x - 1 \leq 0$$

inequality

lagrange mul (λ)

Lagrangian

x λ

$x\lambda$

$$L(x,\lambda) = x^2 - \lambda(x-1)$$

① $\dfrac{\partial L}{\partial x} = 0 \quad \dfrac{\partial L}{\partial \lambda} = 0$

② $x - 1 \leq 0$

③ $\lambda \geq 0$

④ $\lambda(x-1) = 0$

Example

16 July 2023 14:52

$$\boxed{\begin{array}{l} f(x,y) = x^2 + y^2 \quad \underset{x,y}{\text{minimize}} \\ \text{subject to} \quad x+y-1 \le 0 \end{array}}$$

lagrangian

$$\frac{\partial L}{\partial \lambda} = x^2 + y^2 - \lambda x - \lambda y + \lambda$$

$$= -x - y = 0$$

$$-x = y \qquad \boxed{0 \le 0}$$

$$0.5 + 0.5 - 1$$

$$L(x,y,\lambda) = x^2 + y^2 - \lambda(x+y-1)$$

$$\frac{\partial L}{\partial x} = 0 \qquad \frac{\partial L}{\partial y} = 0 \qquad \frac{\partial L}{\partial \lambda} = 0$$

② $x+y-1 \le 0$ ✓

③ $\lambda \ge 0$ ✓

4) $\lambda(x+y-1) = 0$ ✓ $\;>0$

$$2x - \lambda = 0 \qquad 2y - \lambda = 0 \qquad -x - y + 1 = 0$$

$$x - y$$

$$\boxed{x + y = 1}$$

$$\lambda = 2x \qquad \lambda = 2y$$

$$x = \frac{\lambda}{2} \qquad y = \frac{\lambda}{2}$$

$$\frac{\lambda}{2} + \frac{\lambda}{2} = 1 \qquad \boxed{\lambda = 1}$$

$\to 0.5$ $\quad \to 0.5$

$$\boxed{x = 0.5 \quad y = 0.5 \quad \lambda = 1} \to \text{KKT} ✓$$

$$\to \boxed{x = 0.5 \quad y = 0.5}$$

Concept of Duality

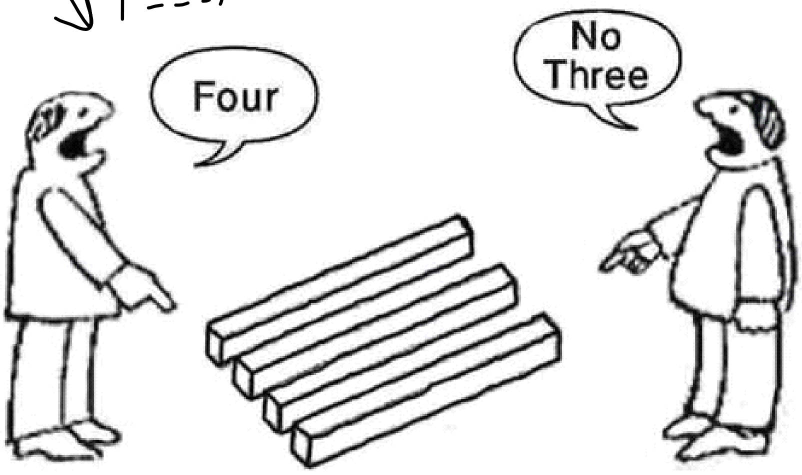
16 July 2023 08:46

eqn → optimize ⟷ dual problem soln

↑ primal ↑ easier to solve

The duality principle is fundamental in optimization theory. It provides a powerful tool for solving difficult or complex optimization problems by transforming them into simpler or easier-to-solve problems. The solution to the dual problem provides a lower bound on the solution of the primal problem. If strong duality holds (i.e., the optimal values of the primal and dual problems are equal), then solving the dual problem can directly give the solution to the primal problem.

→ perspective



The primal problem is the original optimization problem that you are trying to solve. It involves finding the minimum or maximum of a particular objective function, subject to certain constraints.

The dual problem is a related optimization problem that is derived from the primal problem. It provides a lower or upper bound on the solution to the primal problem.

↓

SVM (w, b)

derive → primal → dual

SVM Dual Problem

15 July 2023 23:58

x_1 x_2 x_3 y → w, b train.

primal

Hard margin SVM

The primal form of hard margin SVM is given by:

$$\begin{aligned} &\text{minimize}_{\mathbf{w},b} \quad \frac{1}{2}\|\mathbf{w}\|^2 \\ &\text{subject to} \quad y_i(\mathbf{w}\cdot\mathbf{x}_i + b) \geq 1, \quad i = 1,\ldots,n \end{aligned}$$

w, b → primal variable

$$\begin{aligned} &\text{maximize}_{a} \quad \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j a_i a_j (\mathbf{x}_i\cdot\mathbf{x}_j) \\ &\text{subject to} \quad \boxed{a_i \geq 0,} \quad i = 1,\ldots,n \\ &\qquad \sum_{i=1}^{n} a_i y_i = 0 \end{aligned}$$

w, b

dual variable

→ dual form Hard marg

Hard margin → soft margin

Soft margin SVM

The primal form of soft margin SVM is given by:

$$\begin{aligned} &\text{minimize}_{\mathbf{w},b,\xi} \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i=1}^{n}\xi_i \\ &\text{subject to} \quad y_i(\mathbf{w}\cdot\mathbf{x}_i - b) \geq 1 - \xi_i, \quad i = 1,\ldots,n \\ &\qquad \xi_i \geq 0, \quad i = 1,\ldots,n \end{aligned}$$

$$\begin{aligned} &\text{maximize}_{\mathbf{a}} \quad \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j a_i a_j (\mathbf{x}_i\cdot\mathbf{x}_j) \\ &\text{subject to} \quad \boxed{0 \leq a_i \leq C,} \quad i = 1,\ldots,n \\ &\qquad \sum_{i=1}^{n} a_i y_i = 0 \end{aligned}$$

dual

Dual Problem Derivation

16 July 2023 01:14

n rows ↓ $x_1 \; x_2 \; x_3 \ldots x_m \mid y$ (m cols)

Primal form

$$\underset{w,b}{\operatorname{argmin}} \left[\frac{\|w\|^2}{2} \right]$$

such that

$$\underbrace{y_i(w^T x_i + b) \geq 1}_{} \; \forall i$$

↳ how many constraints

for every row
↳ 1 constraint

$$L(w,b,\alpha) = \frac{\|w\|^2}{2} - \alpha_1 [y_1(w^T x_1 + b) - 1] - \alpha_2 [y_2(w^T x_2 + b) - 1] - \ldots - \alpha_n [y_n(w^T x_n + b) - 1]$$

$$L(w,b,\alpha_i) = \frac{\|w\|^2}{2} - \sum_{i=1}^{n} \alpha_i [y_i(w^T x_i + b) - 1]$$

$$\frac{\partial L}{\partial w} = 0 \qquad \frac{\partial L}{\partial b} = 0$$

$$= \frac{\|w\|^2}{2} - \sum_{i=1}^{n} \alpha_i y_i w^T x_i \pm \alpha_i y_i b - \alpha_i$$

$$L(w,b,\alpha_i) = \frac{\|w\|^2}{2} - \sum_{i=1}^{n} \alpha_i y_i w . x_i - \sum_{i=1}^{n} \alpha_i y_i b + \sum_{i=1}^{n} \alpha_i$$

$$\frac{\partial L}{\partial w} = \frac{\cancel{2} w}{\cancel{2}} - \sum_{i=1}^{n} \alpha_i y_i x_i = 0$$

$$\boxed{w = \sum_{i=1}^{n} \alpha_i y_i x_1} \rightarrow (1)$$

$$\frac{\partial L}{\partial b} = -\sum_{i=1}^{n} \alpha_i y_i = 0$$

$$\boxed{\sum_{i=1}^{n} \alpha_i y_i = 0} \leftarrow$$

$$L(w,b,\alpha_i) = \frac{\|w\|^2}{2} - \sum_{i=1}^{n} \alpha_i y_i w\cdot x_i - \sum_{i=1}^{n} \alpha_i y_i b + \sum_{i=1}^{n} \alpha_i$$

$$= \frac{1}{2}\left(\sum_{i=1}^{n} \alpha_i y_i x_i\right)\left(\sum_{j=1}^{n} \alpha_j y_j x_j\right) - \left(\sum_{i=1}^{n} \alpha_i y_i x_i\right)\left(\sum_{j=1}^{n} \alpha_j y_j x_j\right) + \sum_{i=1}^{n} \alpha_i$$

$$-\frac{1}{2}\sum_{i=1}^{n} \alpha_i y_i x_i \sum_{j=1}^{n} \alpha_j y_j x_j + \sum_{i=1}^{n} \alpha_i$$

(α_i) $\quad$ ~~w, b~~ x

$$\boxed{\sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)} \rightarrow \text{dual form}$$

$\boxed{\alpha_i \geq 0}$

$$\sum_{i=1}^{n} \alpha_i y_i = 0$$

$\alpha_i \leftarrow w, b$

j i →
j i →
j i →
j →

$$\boxed{\text{maximize}}_{\alpha_i} \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j \alpha_i \alpha_j (x_i x_j) \rightarrow \text{dot product}$$

min → max $\qquad \alpha_i \rightarrow$ # rows $\qquad$ n lagran

$$\left\{\begin{matrix} \text{min} \rightarrow \text{max} \\ \text{max} \rightarrow \text{min} \end{matrix}\right\}$$

Observations

15 July 2023 23:59

1) easy to solve
2) kernel friendly

The primal form of hard margin SVM is given by:

$$\text{minimize}_{\mathbf{w},b} \quad \frac{1}{2}\|\mathbf{w}\|^2$$
$$\text{subject to} \quad y_i(\mathbf{w}\cdot\mathbf{x}_i - b) \geq 1, \quad i = 1,\ldots,n$$

1. Alpha_i>0 only for support vectors -> the equation is not as dangerous as it seems
2. Dot product

$\alpha_i \alpha_j$

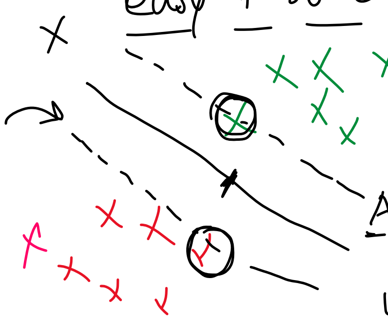
$x\ y \to$ training data

$$\text{maximize}_a \quad \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j a_i a_j (\mathbf{x}_i\cdot\mathbf{x}_j)$$
$$\text{subject to} \quad a_i \geq 0, \quad i = 1,\ldots,n$$
$$\sum_{i=1}^{n} a_i y_i = 0$$

kernel trick

$$\boxed{w = \sum_{i=1}^{n} \alpha_i y_i x_i}$$

easy to solve



$Ax+By+C=0$

$\downarrow$

$wx+b=0$

$\boxed{w} = \boxed{\alpha_1 y_1 x_1} + \alpha_2 y_2 x_2 + \boxed{\alpha_3 y_3 x_3} + \ldots + \alpha_n y_n x_n$

$\boxed{x_2\ y_2}$

$\boxed{w \to \text{support vectors}}$

$\boxed{\alpha \geq 0}$

SV: $\alpha > 0$

Not SV: $\alpha = 0$

for support vector $\alpha > 0$

Non support vector $\alpha = 0$

computation

multiply 2 support vectors

$$\text{maximize}_a \quad \sum_{i=1}^{n} a_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j a_i a_j (\mathbf{x}_i\cdot\mathbf{x}_j)$$
$$\text{subject to} \quad a_i \geq 0, \quad i = 1,\ldots,n$$
$$\sum_{i=1}^{n} a_i y_i = 0$$

row1 $\to$ $x_1\ y_1\ \alpha_1$

row2 $\to$ $x_2\ y_2\ \alpha_2$

rown $\to$ $x_n\ y_n\ \alpha_n$

1 lac $\to$ 5, 10, 15

kernel $\to$ next class