# CS461 – RECITATION 09

## MACHINE LEARNING PRINCIPLES

Daize Dong

2025-11-10

# TODAY’S CONTENT

- Support Vector Machines
- Practice
  - Linear Kernel
  - Polynomial Kernel

# SUPPORT VECTOR MACHINE

Data: $(x_i, y_i)$

Parameters:

- Weight & Bias (hyperplane): $w, w_0$ (vectors)

$$f(x) = w^T x + w_0$$

- Weight & Bias (2D slope–intercept form): $a, b$ (scalars)

$$y = ax + b$$

- Lagrange Multipliers: $\alpha_i$
- Kernel Function: $\mathrm{K}(x, x') = \varphi(x)\varphi(x')$

# SUPPORT VECTOR MACHINE

Other Concepts:

- Margin width $\frac{2}{\|w\|}$: distance from hyperplane to each margin $\frac{1}{\|w\|}$.
- Dual Objective:

$$\max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \mathrm{K}(x_i, x_j)$$

$$\text{s.t.} \quad \sum_{i=1}^{n} \alpha_i y_i = 0 \,, \alpha_i > 0$$

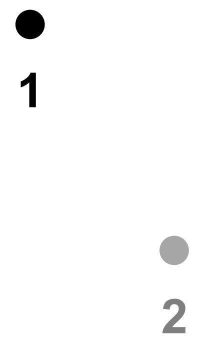
- Only support vectors ($\alpha_i > 0$) shape the model.

# LINEAR KERNEL EXAMPLE

We have 2 data points:

X: (3,1) Y: 1

X: (4,-1) Y: -1




Note that for linear kernels, we have:

$$\mathrm{K}(x, x') = x^T x'$$

$$\Rightarrow \quad \max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j$$

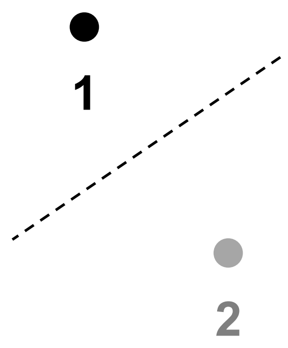
# 1) CALCULATE $\alpha_i$

X: (3,1) Y: 1

X: (4,-1) Y: -1

Due to **geometric symmetry** between the two points, we have

$$\alpha_1 = \alpha_2 \quad (1)$$




So that:

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

$$\Rightarrow (\alpha_1 + \alpha_2) - \frac{1}{2}((9+1)\alpha_1^2 - 2(12-1)\alpha_1\alpha_2 + (16+1)\alpha_2^2)$$

$$\Rightarrow (\alpha_1 + \alpha_2) - \frac{1}{2}(10\alpha_1^2 - 22\alpha_1\alpha_2 + 17\alpha_2^2)$$

$$\xRightarrow[\text{eq. }(1)]{} -\frac{5}{2}\alpha_1^2 + 2\alpha_1$$

# 1) CALCULATE $\alpha_i$

X: (3,1) Y: 1

X: (4,-1) Y: -1

Now the objective turns to:

$$\underset{\alpha_1}{\operatorname{argmax}} -\frac{5}{2}\alpha_1^2 + 2\alpha_1$$

Take the derivative and it is very easy to deduce:

$$\frac{d}{d\alpha_1}\left(-\frac{5}{2}\alpha_1^2 + 2\alpha_1\right) = -5\alpha_1 + 2 = 0$$

So that:

$$\boldsymbol{\alpha_1 = \frac{2}{5}}$$

$$\boldsymbol{\alpha_2 = \alpha_1 = \frac{2}{5}}$$

# 2) SOLVE $w$ AND $w_0$

X: (3,1) Y: 1

X: (4,-1) Y: -1

According to the **stationarity conditions** of the KKT, $w$ satisfy:

$$\alpha_1 = \frac{2}{5}$$

$$\alpha_2 = \frac{2}{5}$$

$$\begin{aligned} w &= \sum_{i=1}^{2} \alpha_i y_i x_i \\ &= \frac{2}{5} \cdot 1 \cdot \begin{pmatrix} 3 \\ 1 \end{pmatrix} + \frac{2}{5} \cdot (-1) \cdot \begin{pmatrix} 4 \\ -1 \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{-2/5} \\ \mathbf{4/5} \end{pmatrix} \end{aligned}$$

Then we have:

$$w_0 = \underline{y_1 - w^T x_1} = 1 - \begin{pmatrix} -2/5 \\ 4/5 \end{pmatrix}^T \begin{pmatrix} 3 \\ 1 \end{pmatrix} = \frac{\mathbf{7}}{\mathbf{5}}$$

You can use any support vector here as $x_i, y_i$

# 3) NORMAL LINEAR EQUATION

This means we need to convert the hyperplane into normal format:

$$f(x) = w^T x + w_0 \quad \rightarrow \quad y = ax + b$$

$$w = \begin{pmatrix} -2/5 \\ 4/5 \end{pmatrix}$$

$$w_0 = \frac{7}{5}$$

It is very easy, just replace $x$ with $\begin{pmatrix} x \\ y \end{pmatrix}$ and let $f(x) = 0$:

$$\begin{pmatrix} -2/5 \\ 4/5 \end{pmatrix}^T \begin{pmatrix} x \\ y \end{pmatrix} + \frac{7}{5} = 0$$

So that:

$$y = \frac{1}{2}x - \frac{7}{4}$$

# 4) MARGIN SIZE $\frac{1}{\|w\|}$

Just take $w$ in:

$$w = \begin{pmatrix} -2/5 \\ 4/5 \end{pmatrix}$$

$$\frac{1}{\|w\|} = \frac{1}{\sqrt{\left(-\frac{2}{5}\right)^2 + \left(\frac{4}{5}\right)^2}}$$

$$= \frac{\sqrt{\mathbf{5}}}{\mathbf{2}}$$

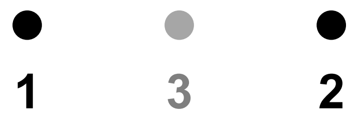
# POLYNOMIAL KERNEL EXAMPLE

We have 3 data points:

X: (-1,0) Y: 1

X: (1,0) Y: 1

X: (0,0) Y: -1




Assume we use 2-order polynomial kernel, we have:

$$\mathrm{K}(x, x') = (x^T x' + 1)^2$$

$$\Rightarrow \quad \max_{\alpha} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j (x_i^T x_j + 1)^2$$

# 1) CALCULATE $\alpha_i$

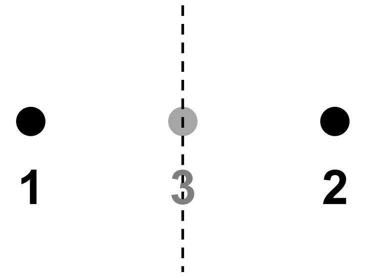
X: (-1,0) Y: 1

X: (1,0) Y: 1

X: (0,0) Y: -1

Due to **geometric symmetry** between the first two points, we have

$$\alpha_1 = \alpha_2 \quad (1)$$




WHY?

Because we only have 3 points.

Meanwhile we have:

$$\sum_{i=1}^{n} \alpha_i y_i = 0 \quad \Rightarrow \quad \alpha_1 + \alpha_2 - \alpha_3 = 0$$

$$\Rightarrow \quad \alpha_3 = 2\alpha_1 \quad (2)$$

# 1) CALCULATE $\alpha_i$

X: (-1,0) Y: 1

X: (1,0) Y: 1

X: (0,0) Y: -1

So that:

$$\sum_{i=1}^{3} \alpha_i - \frac{1}{2} \sum_{i=1}^{3} \sum_{j=1}^{3} \alpha_i \alpha_j y_i y_j (x_i^T x_j + 1)^2$$

$\alpha_2 = \alpha_1$
$\alpha_3 = 2\alpha_1$

$$\Rightarrow \; (\alpha_1 + \alpha_2 + \alpha_3) - \frac{1}{2}((1 + 0 + 1)\alpha_1^2 + \cdots + (0 + 0 + 1)\alpha_3^2)$$

$$\Rightarrow \; \dots \dots$$

$$\xRightarrow[\text{eq. (1) and (2)}]{} \; -2\alpha_1^2 + 4\alpha_1$$

# 1) CALCULATE $\alpha_i$

X: (-1,0) Y: 1

X: (1,0) Y: 1

X: (0,0) Y: -1

Now the objective turns to:

$$\underset{\alpha_1}{\operatorname{argmax}} -2\alpha_1^2 + 4\alpha_1$$

Take the derivative and it is very easy to deduce:

$$\frac{\mathrm{d}}{\mathrm{d}\alpha_1}(-2\alpha_1^2 + 4\alpha_1) = -4\alpha_1 + 4 = 0$$

So that:

$$\boldsymbol{\alpha_1 = 1}$$
$$\boldsymbol{\alpha_2 = \alpha_1 = 1}$$
$$\boldsymbol{\alpha_3 = 2\alpha_1 = 2}$$

# 2) SOLVE $w_0$

X: (-1,0) Y: 1

X: (1,0) Y: 1

X: (0,0) Y: -1

Just bring in the equations:

$$w_0 = y_i - \boxed{K(w, x_i)}$$

$$= y_i - \boxed{\sum_j^3 \alpha_j y_j K(x_i, x_j)}$$

$$= y_1 - \sum_j^3 \alpha_j y_j K(x_1, x_j)$$

$$= y_1 - \left((x_1{}^T x_1 + 1)^2 + (x_1{}^T x_2 + 1)^2 - 2(x_1{}^T x_3 + 1)^2\right)$$

$$= 1 - \left((1+1)^2 + (-1+1)^2 - 2(0+1)^2\right)$$

$$= 1 - (4 + 0 - 2)$$

$$= \mathbf{-1}$$

$$w = \sum_j^3 \alpha_j y_j \varphi(x_j)$$

$$w\varphi(x_i) = \sum_j^3 \alpha_j y_j \varphi(x_j)\varphi(x_i)$$

$$K(w, x_i) = \sum_j^3 \alpha_j y_j K(x_i, x_j)$$

$\alpha_1 = 1$
$\alpha_2 = 1$
$\alpha_3 = 2$

You can use any support vector here as $x_i, y_i$