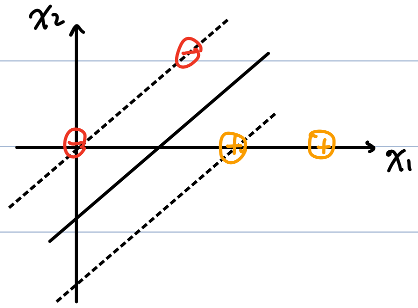
Lec 28 Lagrange Dual Formulation of SVM

Recap: SVM



↳ Fact (optional)

$$E_{out} \approx E_{cross\ validation} \le \frac{\#\text{ support vectors}}{N}$$

↳ Original formulation:

$$\min_{\underline{w}, b} \frac{1}{2} \underline{w}^T \underline{w} \quad \text{s.t.} \quad y_n (\underline{w}^T \underline{x}_n + b) \ge 1, \quad n = 1, \cdots N$$

"primal problem"

1. Lagrange Dual Formulation of SVM (ch 8.2)

↳ Lagrange function:

$$\mathcal{L}(\underline{w}, b, \underline{\alpha}) = \frac{1}{2} w^T w \underbrace{- \sum_{n=1}^{N} \alpha_n (y_n (\underline{w}^T \underline{x}_n + b) - 1)}_{\text{penalty for constraints violation}},$$

for $\alpha_1 \cdots \alpha_N \ge 0$

"dual variables" / "Lagrange Multiplier"

↳ <u>Fact:</u> Primal problem is convex

Mild cond. $\Longrightarrow$ "strong duality" property

i.e. optimal solution to primal problem

$$\Updownarrow$$

$$\underbrace{\max_{\{\alpha_n \geq 0\}} \underbrace{\min_{\underline{w}, b} \mathcal{L}(\underline{w}, b, \underline{\alpha})}_{\text{first}}}_{\text{second}}$$

unconstrained optimization

↳ <u>Steps:</u>

① $\nabla_{\underline{w}} \mathcal{L}(\underline{w}, b, \underline{\alpha}) = 0$

$$\Rightarrow \underline{w} - \sum_{n=1}^{N} \alpha_n y_n \underline{x}_n = 0$$

$$\Rightarrow \underline{w} = \sum_{n=1}^{N} \alpha_n y_n \underline{x}_n$$

$$\frac{\partial \mathcal{L}(\underline{w}, b, \underline{\alpha})}{\partial b} = 0 \Rightarrow \sum_{n=1}^{N} \alpha_n y_n = 0$$

$$\Rightarrow \mathcal{L}(\underline{w}, b, \underline{\alpha}) = \frac{1}{2} \underline{w}^T \underline{w} - \underline{w}^T \underline{w} + \sum_{n=1}^{N} \alpha_n$$

$$= -\frac{1}{2} \underline{w}^T \underline{w} + \sum_{n=1}^{N} \alpha_n$$

$$\therefore \mathcal{L}(\underline{\alpha}) = -\frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m \underline{x}_n^T \underline{x}_n + \sum_{n=1}^{N} \alpha_n$$

↑ remove dependency of $\underline{w}$, b

② $\max \mathcal{L}(\underline{\alpha})$ s.t. $\underline{\alpha}_n \geq 0$, $n = 1, 2, \cdots N$ "Dual problem"

$\sum_{n=1}^{N} \alpha_n y_n = 0$ // Another inner constraints

This is also a QP problem

$$\mathcal{L}(\underline{\alpha}) = \frac{1}{2}\underline{\alpha}^T Q \underline{\alpha} + p^T \underline{\alpha}$$

$$Q = \begin{bmatrix} y_1 y_1 \underline{x}_1^T \underline{x}_1 & y_1 y_2 \underline{x}_1^T \underline{x}_2 & \cdots & y_1 y_N \underline{x}_1^T \underline{x}_N \\ \vdots & & & \\ \vdots & & & \\ y_N y_1 \underline{x}_N^T \underline{x}_1 & & \cdots & y_N y_N \underline{x}_N^T \underline{x}_N \end{bmatrix}, \in \mathbb{R}^{N \times N}$$

$$\underline{p} = \begin{bmatrix} -1 \\ \vdots \\ -1 \end{bmatrix}$$

constraints: linear

↳ if $\underline{\alpha}^*$ is an optimal solution to the dual problem, then

$$\underline{w}^* = \sum_{n=1}^{N} \alpha_n^* y_n \underline{x}_n$$

↳. How to find b ?

- One the KKT conditions
- At optimum, we must have:

$$\alpha_n (y_n (\underline{w}^T \underline{x}_n + b) - 1) = 0 \quad , \forall n$$

(penlty = 0)

$\therefore$ either $\alpha_n = 0$ or $y_n (\underline{w}^T \underline{x} + b) - 1 = 0$

- If $\alpha_n \neq 0$, then $\underline{x}_n$ is a point on the margin

  i.e. is a support vector

- If $\alpha_n = 0$, then $\underline{x}_n$ does not affect $\underline{w}$.

  and $(\underline{x}_n, y_n)$ can be removed from $\mathcal{D}$ w/out changing the sol.

- To find b^*, pick any support vector $(\underline{x}_s, y_s)$ s.t

$$y_s(\underline{w}^T \underline{x}_s + b) = 1$$

$\because$ $y_s \in \{+1, -1\}$ $\quad\therefore$ $\frac{1}{y_s} = y_s$

$$\therefore \quad b = y_s - \underline{w}^T \underline{x}_s = y_s - \sum_{n=1}^{N} y_n \alpha_n \underline{x}_n^T \underline{x}_s$$