SVM

1 Brief introduction

Assume the the data is linearly separable. We can convert our question into math equation.

$$
\begin{aligned}
\max_{w} \quad & margin(w) \\
\text{subject to} \quad & every \quad y_n w^T x_n \\
& margin(w) = \min_{n=1\ldots N} distance(x_n, w)
\end{aligned}
\tag{1}
$$

Next, we would roughly mention some important techniques used in this algorithm.

The process would be like that :

method to compute(quadratic programming, lagrange multipliers) $\Rightarrow$

selection of kernel $\Rightarrow$

trade-off between large margin and margin violation(soft SVM) $\Rightarrow$

method to select the C (the trade-off) and the γ (the parameter when using gaussian kernel)$\Rightarrow$

Empirical results(of the expenditure analysis)

2 method to compute

1. quadratic programming :

$$\begin{aligned} \min_{u} \quad & \frac{1}{2}u^TQu + p^Tu \\ \text{subject to} \quad & a_m^Tu \geq c_m \\ & for \quad m = 1, 2, ..., M \end{aligned} \tag{2}$$

where $u = [b, w]^T; Q = \begin{bmatrix} 0 & 0_d^T \\ 0_d & I_d \end{bmatrix}; \quad p = 0_{d+1}; \quad a_n^T = y_n[1, x_n^T]; \quad c_n = 1; \quad M = N$

2. lagrange multiplier :

original question :

$$\begin{aligned} \min_{b,w} \quad & \frac{1}{2}w^Tw \\ \text{subject to} \quad & y_n(w^Tz_n + b) \geq 1 \\ & for \quad n = 1, 2, ..., N \end{aligned} \tag{3}$$

transfer it to lagrange function:

$$\mathbf{L}(b, w, \alpha) = \frac{1}{2}w^Tw + \sum_{n=1}^{N} \alpha_n(1 - y_n(w^Tz_n + b))$$

3. dual problem :(goal : transfer question from d + 1 variables to N variables)

(strong duality) :

$$\min_{b,w}(\max_{\alpha_n \geq 0} \mathbf{L}(b, w, \alpha)) = \max_{\alpha_n \geq 0}(\min_{b,w} \mathbf{L}(b, w, \alpha))$$

4. KKT :

- $y_n(w^Tz_n + b) \geq 1$
- $\alpha_n \geq 0$
- $\sum y_n\alpha_n = 0; w = \sum \alpha_n y_n z_n$
- $\alpha_n(1 - y_n(w^Tz_n + b)) = 0 \quad (complimentary \quad slackness)$

Note : the "support vectors(z_n, y_n) satisfies $\alpha_n > 0$"

3 selection of kernel

Note : the goal of using kernel SVM is to avoid the dependence of dimension(convert it to dependence on N) and predict with SV only

1. linear kernel :

- $K(x, x') = x^T x'$
- very explainable(w, and SV)
- cons : data not always separable
- we should always begin with it first

2. Gaussian kernel :

   - $K(x, x') = exp(-\gamma||x - x'||^2)$
   - achieve large margin in infinite dimension
   - more powerful than linear and polynomial, but sometimes too powerful
   - need to carefully choose the γ

4 trade-off between large margin and margin violation(soft SVM)

1. margin violation η_n
2. penelized with violation C

$$\begin{aligned} \min_{b,w,\eta} \quad & \frac{1}{2}w^T w + C\sum_{n=1}^{N}\eta_n \\ \text{subject to} \quad & y_n(w^T z_n + b) \geq 1 - \eta_n, \quad where \quad \eta_n \geq 0 \end{aligned} \tag{4}$$

3. physical meaning of α_n

   - complimentary salckness

$$\begin{aligned} & \alpha_n(1 - \eta_n - y_n(w^T z_n + b)) = 0 \\ & (C - \alpha_n)\eta = 0 \end{aligned} \tag{5}$$

   - $\alpha = 0$:
     no SV
     $\eta = 0$
   - $0 < \alpha < C$:
     free SV
     $\eta = 0$
   - $\alpha = C$:
     bounded SV
     $\eta =$ violation amounts

5 method to select the C (the trade-off) and the γ (the parameter when using gaussian kernel)

Note : using cross validation to choose the proper parameters

1. C : the penalized parameter
   - C large : less tolerance of noice $\Rightarrow$ overfits
   - C small : want large margin
2. γ (Gaussian):
   - γ large : sharp Gaussian $\Rightarrow$ overfits

6 Empirical results(of the expenditure analysis)

the code for expenditure analysis:(since I didn't do proper normalization, the results isn't as good as the one provided by our team member 張雅晴)

https://github.com/Raymonna/numerical-computation-p/blob/master/Untitled8.ipynb