quick doc on SVM

Jun Lu

Computer Science, EPFL

jun.lu.locky@gmail.com

1 Note

This documents serves as a quick document for implementation of SVM and might be later a survey on SVM of pairwise feature for SVM.

2 Classification

Let $\{(\boldsymbol{x}_n, y_n)\}_{n=1}^{N}$ be a training sample, normally we have $y_n = \text{sign}(\boldsymbol{w}^T \boldsymbol{x}_n + b)$, where $\boldsymbol{w}, b$ is the weight variable and bias variable we want to learn. In other formulation, the bias variable can also be referred as w_0.

2.1 Hard-margin Primal formulation

Due to trivial derivation, we have

$$\min_{\boldsymbol{w},\boldsymbol{b}} \frac{1}{2} \boldsymbol{w}^T \boldsymbol{w}$$
$$\text{s.t. } y_n(\boldsymbol{w}^T \boldsymbol{x}_n + b) \geq 1, \text{for } n = 1, 2, \ldots, N. \tag{1}$$

2.2 Hard-margin Dual formulation

With Lagrange multiplier α_n and some non-linear function $\boldsymbol{z}_n = \Phi(\boldsymbol{x}_n)$, we have the objective function,

$$\text{L}(\boldsymbol{w}, b, \boldsymbol{\alpha}) = \frac{1}{2} \boldsymbol{w}^T \boldsymbol{w} + \sum_{n=1}^{N} \alpha_n (1 - y_n(\boldsymbol{w}^T \boldsymbol{z}_n + b)). \tag{2}$$

With trivial derivation, we have

$$\min_{\boldsymbol{\alpha}} \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} \alpha_n \alpha_m y_n y_m \boldsymbol{z}_n^T \boldsymbol{z}_m - \sum_{n=1}^{N} \alpha_n$$
$$\text{s.t. } \sum_{n=1}^{N} y_n \alpha_n = 0;$$
$$\alpha_n \geq 0, \text{for } n = 1, 2, \ldots, N. \tag{3}$$

2.3 Kernel hard-margin Dual formulation

Let $q_{n,m} = y_n y_m \boldsymbol{z}_n^T \boldsymbol{z}_m$. What we want is make the calculation of $\boldsymbol{z}_n^T \boldsymbol{z}_m$ faster than $O(D)$, where D is the dimension of sample or transferred feature. Trivial derivation would give $q_{n,m} = y_n y_m K(\boldsymbol{x}_n, \boldsymbol{x}_m)$.

2.4 Soft-margin Primal formulation

Give up some noisy samples

$$\min_{\boldsymbol{w},\boldsymbol{b}} \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C \cdot \sum_{n=1}^{N} \mathbb{1}\{y_n \neq \text{sign}(\boldsymbol{w}^T\boldsymbol{z}_n + b)\}$$

$$\text{s.t. } y_n(\boldsymbol{w}^T\boldsymbol{x}_n + b) \geq 1, \text{for correct n,}$$

$$y_n(\boldsymbol{w}^T\boldsymbol{x}_n + b) \geq -\infty, \text{for incorrect n,}$$

(4)

Equivalently, we have

$$\min_{\boldsymbol{w},\boldsymbol{b}} \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C \cdot \sum_{n=1}^{N} \mathbb{1}\{y_n \neq \text{sign}(\boldsymbol{w}^T\boldsymbol{z}_n + b)\}$$

$$\text{s.t. } y_n(\boldsymbol{w}^T\boldsymbol{x}_n + b) \geq 1 - \infty \cdot \mathbb{1}\{y_n \neq \text{sign}(\boldsymbol{w}^T\boldsymbol{z}_n + b)\}, \text{for all n,}$$

(5)

2.5 Soft-margin Dual formulation

With Lagrange multipliers α_n and β_n, let $\xi_n = \mathbb{1}\{y_n \neq \text{sign}(\boldsymbol{w}^T\boldsymbol{z}_n + b)\}$, , we have the objective function,

$$\text{L}(\boldsymbol{w}, b, \boldsymbol{\xi}, \boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C \cdot \sum_{n=1}^{N} \xi_n$$

$$+ \sum_{n=1}^{N} \alpha_n \cdot (1 - \xi_n - y_n(\boldsymbol{w}^T\boldsymbol{z}_n + b)) \sum_{n=1}^{N} + \beta_n \cdot (-\xi_n).$$

(6)

With trivial derivation, we have

$$\min_{\boldsymbol{\alpha}} \frac{1}{2}\sum_{n=1}^{N}\sum_{m=1}^{N} \alpha_n\alpha_m y_n y_m \boldsymbol{z}_n^T\boldsymbol{z}_m - \sum_{n=1}^{N} \alpha_n$$

$$\text{s.t. } \sum_{n=1}^{N} y_n\alpha_n = 0;$$

$$0 \leq \alpha_n \leq C, \text{for } n = 1, 2, \ldots, N.$$

(7)

3 Regression

Some other references introduced support vector regression (SVR), the basic idea is similar to support vector classification [1]. We here only introduce the soft-margin support vector regression.

3.1 Soft-margin SVR primal formulation

Due to trivial derivation, we have

$$\min_{\boldsymbol{w},\boldsymbol{b}} \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C \cdot \sum_{n=1}^{N} (\xi_n + \xi_n^\star)$$

$$\text{s.t. } y_n(\boldsymbol{w}^T\boldsymbol{z}_n + b) \leq \epsilon + \xi_n,$$

$$y_n(\boldsymbol{w}^T\boldsymbol{z}_n + b) \leq \epsilon + \xi_n^\star,$$

$$\xi_n, \xi_n^\star \geq 0, n = 1, 2, \ldots, N.$$

(8)

3.2 Soft-margin SVR dual formulation

With trivial derivation, we have

$$\min_{\boldsymbol{\alpha}} \frac{1}{2} \sum_{n=1}^{N} \sum_{m=1}^{N} (\alpha_n - \alpha_n^\star)(\alpha_m - \alpha_m^\star) \boldsymbol{z}_n^T \boldsymbol{z}_m - \epsilon \sum_{n=1}^{N} (\alpha_n + \alpha_n^\star) + \sum_{n=1}^{N} y_n (\alpha_n - \alpha_n^\star)$$

$$\text{s.t.} \sum_{n=1}^{N} (\alpha_n - \alpha_n^\star) = 0;$$

$$0 \leq \alpha_n, \alpha_n^\star \leq C, \text{for } n = 1, 2, \dots, N. \tag{9}$$

References

[1] Alex J Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222, 2004.