CS419M

Lecture 10: Soft SVM and Dual Optimization

February 15, 2023

Lecturer: Abir De *Scribe: Group 20*

1 Revisiting the Optimisation Problem

- In the previous lectures we discussed the SVM for seperable cases and non-sperable cases. We introduced the concept of slackness/ relaxation parameter $\xi(x_i,y_i)$ for non seperabele cases.

$$\boldsymbol{y}_i(\boldsymbol{w}^T\boldsymbol{x}_i b) \geq 1 - \xi(\boldsymbol{x}_i, \boldsymbol{y}_i) \tag{1}$$

where $\xi_{x,y} \geq 0$.
To optimize the function we need to **minimize** $\xi_{x,y}$. This gives us:

$$\min_{\boldsymbol{w},\, b,\, \xi_{(\boldsymbol{x}_i, y_i)}} \frac{1}{2}||\boldsymbol{w}||^2 + C\sum_{i \in D} \xi_{(x_i, y_i)}$$

- Convex Optimisation

$$\min_{\theta} f(\theta) \tag{2}$$

such that

$$g(\theta) \leq 0$$

Using Lagrange Multiplier we convert this into a dual optimization problem :

$$\max_{\lambda} \min_{\theta} f(\theta) + \lambda^{\intercal} g(\theta) \tag{3}$$

When f(θ) and g(θ)both are strictly convex, the above two equations are exactly equivalent. We get λ^*,θ^* as optimal solutions then by Slater's condition

$$(\lambda^*)^T \boldsymbol{g}(\theta^*) = 0 \tag{4}$$

2 Continuing with the Optimisation Problem

If the optimization problem is

$$\min_{\theta} f(\theta) \tag{5}$$

such that

$$g(\theta) = 0$$

The dual of this is:

$$\max_{\lambda} \min_{\theta} f(\theta) + \lambda^{\intercal} g(\theta) \tag{6}$$

with no constraints on λ
(Reason: sign of λ discussed below)

Now

$$g(\theta) = 0 \Rightarrow g(\theta) \leq 0, g(\theta) \geq 0 \tag{7}$$

Using previous results we can write

$$\max_{\lambda_1, \lambda_2 \geq 0} \min_{\theta}\ f(\theta) + \boldsymbol{\lambda_1}^{\intercal}(-g(\theta)) + \boldsymbol{\lambda_2}^{\intercal} g(\theta) \tag{8}$$

$$=> \max_{\lambda_1, \lambda_2 \geq 0} \min_{\theta}\ f(\theta) - \boldsymbol{\lambda_1}^{\intercal} g(\theta) + \boldsymbol{\lambda_2}^{\intercal} g(\theta) \tag{9}$$

$$=> \max_{\lambda_1, \lambda_2 \geq 0} \min_{\theta}\ f(\theta) + \lambda^{\intercal} g(\theta) \tag{10}$$

$$\lambda = \lambda_2 - \lambda_1$$

Note that λ can have any sign as both λ_1 and λ_2 are non-negative

2.1 Objective Function

- $$(\omega^*, b^*, \xi^*) = \arg\min_{\omega,b,\xi_i} \frac{1}{2}\,||\omega||^2 + C\sum_{i=1}^{n}\xi_i \tag{11}$$

$$y_i(\omega^\intercal\phi(x_i) + b) \geq 1 - \xi_i \quad \forall\, i = 1, 2...., n$$
$$\xi_i \geq 0 \quad \forall\, i = 1, 2...., n$$

so there are 2n constants overall

Instead of 2n constraints,we can do something better:

$$\xi \geq 0, \forall i \in 1, .., n$$

So we can put all those ξ=0 where $\xi < 0$ while applying the gradient descent algorithm to minimise L.

- Framing our Non-Separable SVM into the previous optimization problem

$$SVM :- \min_{\omega,b,\xi_i} \frac{1}{2}\,||\omega||^2 + C\sum_{i=1}^{n}\xi_i \longleftarrow f(\theta) \tag{12}$$

$$1 - \xi_i - y_i(\omega^\intercal\phi(x_i) + b) \leq 0 \quad \longleftarrow g(\theta) \leq 0$$
$$-\xi_i \leq 0 \longleftarrow g(\theta) \leq\ 0$$

Using Lagrange multipliers we get the following optimization problem

$$\mathcal{L}(\omega, b, \xi_i, \alpha_i, \mu_i)\ = \frac{1}{2}\,||\omega||^2 + C\sum_{i=1}^{n}\xi_i\ + \sum_{i=1}^{n}\alpha_i(1 - \xi_i - y_i(\omega^\intercal\phi(x_i) + b)) + \sum_{i=1}^{n}\mu_i(-\xi_i) \tag{13}$$

So our objective becomes

$$\max_{\alpha\geq 0,\mu\geq 0}\ \min_{\omega,b,\xi}\ \mathcal{L}(\omega, b, \xi_i, \alpha_i, \mu_i) \tag{14}$$

We call

$$\min_{\omega,b,\xi}\ \mathcal{L}(\omega, b, \xi_i, \alpha_i, \mu_i)$$

as the Lagrangian dual function g(α, β).

- First order optimality conditions on g(α,β) give us the following

$$\frac{\partial L}{\partial \omega} = 0 \Rightarrow \omega^* = \sum_{i=1}^{n} \alpha_i y_i x_i \tag{15}$$

$$\frac{\partial L}{\partial \boldsymbol{b}} = 0 \Rightarrow \sum_{i=1}^{n} \alpha_i y_i = 0 \tag{16}$$

$$\frac{\partial L}{\partial \xi_i} = 0 \Rightarrow \alpha_i + \mu_i = C \tag{17}$$

Substituting in the main equation

$$\therefore L = \frac{1}{2}\,||\omega||^2 - \sum \alpha_i\, y_i\, \omega^\intercal x_i \tag{18}$$

substituting $\omega = \sum \alpha_i y_i x_i$

$$\max_{\alpha,\mu} \sum \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j\, y_i\, y_j\, \mathbf{x_i}^\intercal\, x_j \tag{19}$$

$$and\ \ \alpha_i + \mu_i = c\ ,\ \ \sum \alpha_i y_i = 0, \forall i \in D \tag{20}$$

We also see that our objective function is quadratic in $\boldsymbol{w}$ and constraints are linear in $\boldsymbol{w}, \xi_i$. Therefore, all functions are strictly convex , therefore, by Slater's condition, at the optimum

$$\mu_i^* \xi_i^* = 0 \tag{21}$$

$$\alpha_i^*\,(1 - \xi_i^* - y_i(\omega^{*T} x_i + b)) = 0 \tag{22}$$

2.2 Inferences

- So for 'good' points (correctly classified)

$$\xi_i = 0$$

$$1 - y_i(\omega^T x_i + b) < 0$$

$$\Rightarrow \alpha_i = 0$$

- For 'bad' points (misclassified or inside the band)

$$\xi_i > 0$$

$$\mu_i \xi_i = 0$$

$$\Rightarrow \mu_i = 0$$

$$\text{Since } \alpha_i + \mu_i = C \quad \text{we get} \quad \alpha_i = C$$

- For points on the hyperplane

$$\xi_i = 0$$

$$\mu_i = uncertain$$

$$\therefore 0 < \alpha_i < C$$

2.3 Algorithm for finding Optimum Values

First, choose $(\omega^0, b^0, \xi_i^0, \alpha^0 \geq 0, \mu^0 \geq 0) \sim$ Random ()

$$\omega^{k+1} = \omega^k - \nabla_\omega L_{\omega=\omega^k} \tag{23}$$

$$\boldsymbol{b}^{k+1} = \boldsymbol{b}^k - \nabla_{\boldsymbol{b}} L_{\boldsymbol{b}=\boldsymbol{b}^k} \tag{24}$$

$$\xi^{k+1} = \xi^k - \nabla_\xi L_{\xi=\xi^k} \tag{25}$$

$$\alpha^{k+1} = \alpha^k + \nabla_\alpha L_{\alpha=\alpha^k} \tag{26}$$

$$\mu^{k+1} = \mu^k + \nabla_\mu L_{\mu=\mu^k} \tag{27}$$

$$\alpha^{k+1} = ReLU(\alpha^{k+1}) \tag{28}$$

$$\mu^{k+1} = ReLU(\mu^{k+1}) \tag{29}$$

Note that equations (26) and (27) have a '+' sign as we want to maximize w.r.t. $\alpha \; and \; \mu$
Now compute further iterations putting k $\leftarrow$ k+1

3 Brief on Slater's condition

Slater's condition (or Slater condition) is a sufficient condition for strong duality to hold for a convex optimization problem, named after Morton L. Slater. Informally, Slater's condition states that the feasible region must have an interior point.

Consider the following optimization problem:

$$\text{Minimize} \;\; f_0(x)$$

$$\text{subject to:}$$

$$f_i(x) \leq 0, i = 1, \ldots, m$$

where $f_i(x)$ are convex functions. In words, Slater's condition for convex programming states that strong duality holds if there exists an x^* such that x^* is strictly feasible (i.e. all constraints are satisfied and the nonlinear constraints are satisfied with strict inequalities).

Mathematically, Slater's condition states that strong duality holds if there exists an $x^* \in \operatorname{relint}(D)$ (where relint denotes the relative interior of the convex set $D := \cap_{i=0}^{m} \operatorname{dom}(f_i)$) such that

$$f_i(x^*) < 0, i = 1, \ldots, m \text{ (the convex, nonlinear constraints)}$$

Also at optimum x^*, $LangrangianMultiplier*$f$_i$(x*)=0

For any convex set $C \subseteq \mathbb{R}^n$ the relative interior is defined as:

$$\operatorname{relint}(C) := \{x \in C : \text{ for all } y \in C, \text{ there exists some } \lambda > 1 \text{ such that } \lambda x + (1-\lambda)y \in C\}.$$

4 Brief Discussion of Concavity/Convexity of Dual Problem

4.1 Maximising Objective function

$$\max_{\lambda} \min_{\theta} f(\theta) + \lambda^{\intercal} g(\theta) \tag{30}$$

whatever be the inner problem (convex or concave) the outer problem is always concave.
Hint for Proof:
$\min_\theta f(\theta) + \lambda^{\intercal} g(\theta)$as sum over various θ and calculate it's mean. Now maximise this mean, as we know max of function is concave, Hence the objective function is concave

4.2 Mention of Jensen' Inequality

E(f(x)) > f(E(x)) **for convex functions**.

Proof-
Suppose f is differentiable. The function f is convex if, for any x and y,

$$\frac{f(x)-f(y)}{(x-y)} \leq f'(y)$$

Let x = X and y = E[X].

We can write

f(X) $\leq$ f(E[X]) + (X $-$ E[X])f ' (E[X])

This inequality is true for all X, so we can take expectation on both sides to get

E[f(X)] $\leq$ f(E[X]) + f ' (E[X])E[(X $-$ E[X])] = f(E[X])

5 References

- Brief on Slater Condition- Wikipedia,https://en.wikipedia.org/wiki/Slater
- CS419 Scribes,2023, Lec 9