CS 419M Introduction to Machine Learning **Spring 2021-22**

Lecture 10: Regression Loss Function Duality

Lecturer: Abir De *Scribe: Group 2*

10.1 Recap

10.1.1 Why is Bias Term not Regularised?

For the loss function given as

$$F_D(W) \quad = \quad \Sigma_{i\in D}[l(w^T x_i + b, y_i) + \lambda||w||^2]$$

Bias term b is usually not regularised. This gives a default value of prediction when no data is provided for training the model.

If the loss function is not regularised with respect to bias term, the loss function not stable, As the minimum of the eigenvalue of hessian of loss function is 0. Meaning the loss function is not strictly convex.

$$min\ eig\left(\frac{\partial^2 F}{\partial b^2}\right) = 0 \Rightarrow F(w,b)\ is\ not\ strictly\ convex$$

10.2 Regression Dual Formulation

Optimization model for support vector machine

$$\underset{w,b}{min}\ \ ||w||^2 + C\Sigma_{i\in D}[1-(w^T x_i + b)y_i)]\ ,\ where\ C = \frac{1}{\lambda}$$

can be equivalently written in the form

$$\underset{w,b,\zeta_i\geq 0}{min}\ ||w||^2 + \underset{i\in D}{\Sigma} C\zeta_i$$

$$y_i(w^T x_i + b) \geq 1 - \zeta_i \ \ \forall\ i \in D$$

This problem in constrained space can be converted to unconstrained dual, which is given as

$$\underset{\alpha_i\geq 0,\beta_i\geq 0}{max}\ \underset{w,b,\zeta\geq 0}{min}\ \ ||w||^2 + \underset{i\in D}{\Sigma}\left[C\zeta_i + \alpha_i(1-\zeta_i - y_i(w^T x_i + b)) - \beta_i\zeta_i\right]$$

10.3 Simplifying Dual Equation

Take $F_{w,b,\zeta} = ||w||^2 + \sum\limits_{i \in D} \left[C\zeta_i + \alpha_i(1 - \zeta_i - y_i(w^T x_i + b)) - \beta_i \zeta_i\right]$

Now,

$$\frac{\partial F_{w,b,\zeta}}{\partial w} = 2w - y_i \sum_{i \in D} x_i \alpha_i = 0$$

$$w = \frac{\sum\limits_{i \in D} x_i \alpha_i y_i}{2} \quad .....(i)$$

$$\frac{\partial F_{w,b,\zeta}}{\partial b} = 0 \Rightarrow \sum_{i \in D} \alpha_i y_i = 0 \quad .....(ii)$$

$$\frac{\partial F_{w,b,\zeta}}{\partial \zeta} = 0 \Rightarrow c - \alpha_i - \beta i = 0$$

$$\alpha_i + \beta i = c \quad .....(iii)$$

Putting these relations, above dual equation becomes:

$$\max_{\alpha_i \geq 0} \sum_{i \in D} \alpha_i - \frac{1}{4} \sum_{i \in D, j \in D} \alpha_i \alpha_j x_i^T x_j y_i y_j$$

$\sum\limits_{i \in D} \alpha_i - \frac{1}{4} \sum\limits_{i \in D, j \in D} \alpha_i \alpha_j x_i^T x_j y_i y_j$ is concave. And as $\alpha_i + \beta i = c$ and $\beta_i \geq 0$, so $0 \leq \alpha_i \leq c$

10.4 Applying Constainsts and Conditions

From Slatu's condition, following two conditions must be satisfied:

$$\alpha_i((w^T x_i + b)y_i - 1 + \zeta_i) = 0 \tag{10.1}$$

$$\beta_i \zeta_i = 0 \tag{10.2}$$

Let's consider some cases:

(i) $y_i(w^T x_i + b) > 1$ for some $(x_i, y_i), \alpha_i =?$
Ans. As $\beta_i \geq 0 \;\Rightarrow\; \zeta_i = 0 \;\Rightarrow\; \alpha_i = 0$

(ii) $y_i(w^T x_i + b) < 1$ for some $(x_i, y_i), \alpha_i =?$
Ans. As $\zeta_i > 0$ (from eq^n 10.1) $\;\Rightarrow\; \beta_i = 0$ (from eq^n 10.2) $\Rightarrow\; \alpha_i = c$ (as $\alpha_i + \beta_i = c$)

(iii) $y_i(w^T x_i + b) = 1$ for some $(x_i, y_i), \alpha_i =?$
Ans. As $\alpha_i \zeta_i = 0$ (from eq^n 10.1) $\Rightarrow$ $\zeta_i = 0$ (using eq^n 10.2) & As $\beta_i \geq 0$ $\Rightarrow$ $0 < \alpha \leq c$ (as $\alpha_i + \beta_i = c$)

10.5 Group Details and Individual Contribution

- 200020112 Sahil Khan
- 190040120 Sumeet Ranjan
- 200040038 Chaitanya Langde
- 19B080018 Prateek Kumar