MLT Week-11

Sherry Thomas
21f3001449

Contents

Abstract

This week's curriculum entails a further examination of the perceptron algorithm, followed by a comprehensive exploration of support vector machines (SVM) and subsequently, an elaboration on the concept of soft-margin SVM.

Dual Formulation for Soft-Margin SVM

The primal formulation for Soft-Margin SVM is given by,

$$\min_{w,\epsilon} \frac{1}{2}||w||_2^2 + C\sum_{i=1}^{n}\epsilon_i \quad s.t. \quad (w^Tx_i)y_i + \epsilon_i \geq 1; \quad \epsilon_i \geq 0 \quad \forall i$$

where C is the hyperparameter that is used to balance the trade-off between maximizing the margin and minimizing the number of misclassifications, and ϵ_i is the additional value required to satisy the constraints.

The Lagrangian function $\mathcal{L}(w, \epsilon, \alpha, \beta)$ for our above function is defined as follows:

$$\mathcal{L}(w, \epsilon, \alpha, \beta) = \frac{1}{2}||w||_2^2 + C\sum_{i=1}^{n}\epsilon_i + \sum_{i=1}^{n}\alpha_i(1 - (w^Tx_i)y_i - \epsilon_i) + \sum_{i=1}^{n}\beta(-\epsilon_i)$$

where $\alpha_i \geq 0$ and $\beta_i \geq 0$ $\forall i$.

Dual Formulation

Maximizing the Lagrangian function w.r.t. α and β, and minimizing it w.r.t. w and ϵ, we get,

$$\min_{w,\epsilon}\left[\max_{\alpha\geq 0;\beta\geq 0} \frac{1}{2}||w||_2^2 + C\sum_{i=1}^{n}\epsilon_i + \sum_{i=1}^{n}\alpha_i(1 - (w^Tx_i)y_i - \epsilon_i) + \sum_{i=1}^{n}\beta(-\epsilon_i)\right]$$

The dual of this is given by,

$$\max_{\alpha\geq 0;\beta\geq 0}\left[\min_{w,\epsilon}\frac{1}{2}||w||_2^2 + C\sum_{i=1}^{n}\epsilon_i + \sum_{i=1}^{n}\alpha_i(1-(w^Tx_i)y_i-\epsilon_i) + \sum_{i=1}^{n}\beta(-\epsilon_i)\right]$$

$$\max_{\alpha\geq 0;\beta\geq 0}\left[\min_{w,\epsilon}\mathcal{L}(w,\epsilon,\alpha,\beta)\right] \quad ...[1]$$

Differentiating the above function[1] w.r.t. w while fixing α and β, we get,

$$\frac{d\mathcal{L}}{dw} = 0$$

$$\frac{d}{dw}\frac{1}{2}||w||_2^2 + C\sum_{i=1}^{n}\epsilon_i + \sum_{i=1}^{n}\alpha_i(1-(w^Tx_i)y_i-\epsilon_i) + \sum_{i=1}^{n}\beta(-\epsilon_i) = 0$$

$$w^*_{\alpha,\beta} - \alpha_i x_i y_i = 0$$

$$\therefore w^*_{\alpha,\beta} = \alpha_i x_i y_i \quad ...[2]$$

Differentiating the above function[1] w.r.t. $\epsilon_i \forall i$ while fixing α and β, we get,

$$\frac{\partial\mathcal{L}}{\partial\epsilon_i} = 0$$

$$\frac{\partial}{\partial\epsilon_i}\frac{1}{2}||w||_2^2 + C\sum_{i=1}^{n}\epsilon_i + \sum_{i=1}^{n}\alpha_i(1-(w^Tx_i)y_i-\epsilon_i) + \sum_{i=1}^{n}\beta(-\epsilon_i) = 0$$

$$C - \alpha_i - \beta_i = 0$$

$$\therefore C = \alpha_i + \beta_i \quad ...[3]$$

Substituting the values of w and β from [2] and [3] in [1], we get,

$$\max_{\alpha\geq 0;\beta\geq 0;C=\alpha_i+\beta_i}\left[\frac{1}{2}||\alpha_i x_i y_i||_2^2 + C\sum_{i=1}^{n}\epsilon_i + \sum_{i=1}^{n}\alpha_i(1-((\alpha_i x_i y_i)^T x_i)y_i-\epsilon_i) + \sum_{i=1}^{n}(C-\alpha_i)(-\epsilon_i)\right]$$

$$\max_{\alpha\geq 0;\beta\geq 0;C=\alpha_i+\beta_i}\left[\frac{1}{2}\alpha_i^T x_i^T y_i^T y_i x_i \alpha_i + C\sum_{i=1}^{n}\epsilon_i + \sum_{i=1}^{n}\alpha_i - \alpha_i^T x_i^T y_i^T y_i x_i \alpha_i - \sum_{i=1}^{n}\alpha_i\epsilon_i - C\sum_{i=1}^{n}\epsilon_i + \sum_{i=1}^{n}\alpha_i\epsilon_i\right]$$

$$\max_{\alpha\geq 0;\beta\geq 0;C=\alpha_i+\beta_i}\left[\sum_{i=1}^{n}\alpha_i - \frac{1}{2}\alpha_i^T x_i^T y_i^T y_i x_i \alpha_i\right]$$

$$\therefore \max_{0\leq\alpha\leq C}\left[\sum_{i=1}^{n}\alpha_i - \frac{1}{2}\alpha_i^T x_i^T y_i^T y_i x_i \alpha_i\right]$$

Credits

Professor Arun Rajkumar: The content as well as the notations are from his slides and lecture.