ONLINE MASTERS IN **DATA SCIENCE**

DSC 255 - MACHINE LEARNING FUNDAMENTALS

DUALITY IN LINEAR CLASSIFICATION

SANJOY DASGUPTA, PROFESSOR

UC San Diego
COMPUTER SCIENCE & ENGINEERING
HALICIOĞLU DATA SCIENCE INSTITUTE

Topics We'll Cover

1. Dual form of the Perceptron
2. Dual form of the support vector machine

Dual Form of the Perceptron Solution

Given a training set of points $\left\{\left(x^{(i)}, y^{(i)}\right): i = 1 \dots n\right\}$:

The Perceptron Algorithm

- Initialize $w = 0$ and $b = 0$
- While some training point (x, y) is misclassified:
  - $w = w + yx$
  - $b = b + y$

Dual Form of the Perceptron Solution

Given a training set of points $\{(x^{(i)}, y^{(i)}): i = 1 \dots n\}$:

The Perceptron Algorithm

- Initialize $w = 0$ and $b = 0$
- While some training point (x, y) is misclassified:
  - $w = w + yx$
  - $b = b + y$

The final answer is of the form:

$$w = \sum_i \alpha_i y^{(i)} x^{(i)},$$

where $\alpha_i = \#$ of time an update occurred on point i.

Dual Form of the Perceptron Solution

Given a training set of points $\{(x^{(i)}, y^{(i)}): i = 1 \dots n\}$:

The Perceptron Algorithm

- Initialize $w = 0$ and $b = 0$
- While some training point (x, y) is misclassified:
  - $w = w + yx$
  - $b = b + y$

The final answer is of the form:

$$w = \sum_i \alpha_i y^{(i)} x^{(i)},$$

where $\alpha_i = \#$ of time an update occurred on point i.

Can equivalently represent w by $\alpha = (\alpha_1, \dots, \alpha_n)$.

Dual Form of the Perceptron Algorithm

Perceptron algorithm: primal form

- Initialize $w = 0$ and $b = 0$
- While some training point $(x^{(i)}, y^{(i)})$ is misclassified:
  - $w = w + y^{(i)} x^{(i)}$
  - $b = b + y^{(i)}$

Perceptron algorithm: dual form

- Initialize $\alpha = 0$ and $b = 0$
- While some training point $(x^{(i)}, y^{(i)})$ is misclassified:
  - $\alpha_i = \alpha_i + 1$
  - $b = b + y^{(i)}$

Answer: $w = \sum_i \alpha_i y^{(i)} x^{(i)}$

Hard-margin SVM

- Initialize $\left(x^{(i)}, y^{(i)}\right), \ldots, \left(x^{(n)}, y^{(n)}\right) \in \mathbb{R}^d \times \{-1, +1\}$

$$\text{(PRIMAL)} \qquad \min_{w \in \mathbb{R}^d, b \in \mathbb{R}} \qquad \|w\|^2$$

$$\text{s.t.: } y^{(i)}(w \cdot x^{(i)} + b) \geq 1 \quad \text{for all } i = 1, 2, \ldots, n$$

- This is a **convex optimization problem**:
  - ➢ Convex objective function
  - ➢ Linear constraints

Hard-margin SVM

- Initialize $\left(x^{(i)}, y^{(i)}\right), \dots, \left(x^{(n)}, y^{(n)}\right) \in \mathbb{R}^d \times \{-1, +1\}$

$$\boxed{\begin{array}{c} \text{(PRIMAL)} \qquad \min_{w \in \mathbb{R}^d, b \in \mathbb{R}} \quad \|w\|^2 \\[1em] \text{s.t.: } y^{(i)}\left(w \cdot x^{(i)} + b\right) \geq 1 \quad \text{for all } i = 1, 2, \dots, n \end{array}}$$

- This is a **convex optimization problem**:
  - Convex objective function
  - Linear constraints
- As such, it has a **dual maximization problem**.
- The **primal** and **dual** problems have the same optimum value.

The Dual Program

$$\text{(PRIMAL)} \qquad \min_{w \in \mathbb{R}^d, b \in \mathbb{R}} \qquad \|w\|^2$$

s.t.: $y^{(i)}(w \cdot x^{(i)} + b) \geq 1$ for all $i = 1, 2, \ldots, n$

$$\text{(DUAL)} \qquad \max_{\alpha \in \mathbb{R}^n} \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{ij=1}^{n} \alpha_i \alpha_j y^{(i)} y^{(j)} \left(x^{(i)} \cdot x^{(j)} \right)$$

$$\text{s.t.:} \sum_{i=1}^{n} \alpha_i \, y^{(i)} = 0$$

$$\alpha \geq 0$$

The Dual Program

$$\text{(PRIMAL)} \qquad \min_{w \in \mathbb{R}^d, b \in \mathbb{R}} \qquad \|w\|^2$$

s.t.: $y^{(i)}(w \cdot x^{(i)} + b) \geq 1$ for all $i = 1, 2, \dots, n$

$$\text{(DUAL)} \qquad \max_{\alpha \in \mathbb{R}^n} \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{ij=1}^{n} \alpha_i \alpha_j y^{(i)} y^{(j)} \left(x^{(i)} \cdot x^{(j)}\right)$$

$$\text{s.t.:} \sum_{i=1}^{n} \alpha_i \, y^{(i)} = 0$$

$$\alpha \geq 0$$

Complementary slackness: At optimality, $w = \sum_{i=1}^{n} \alpha_i y^{(i)} x^{(i)}$ and

$$\alpha_i > 0 \Longrightarrow y^{(i)}\left(w \cdot x^{(i)} + b\right) = 1$$

Points $x^{(i)}$ with $\alpha_i > 0$ are **support vectors**.

Dual of Soft-Margin SVM

$$\text{(PRIMAL)} \quad \min_{w\in\mathbb{R}^d, b\in\mathbb{R}, \xi\in\mathbb{R}^n} \quad \|w\|^2 + C\sum_{i=1}^{n}\xi_i$$

$$\text{s.t.: } y^{(i)}\left(w\cdot x^{(i)} + b\right) \geq 1 - \xi_i \quad \text{for all } i = 1, 2, \dots, n$$

$$\xi \geq 0$$

$$\text{(DUAL)} \quad \max_{\alpha\in\mathbb{R}^n} \quad \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{ij=1}^{n}\alpha_i\alpha_j y^{(i)}y^{(j)}\left(x^{(i)}\cdot x^{(j)}\right)$$

$$\text{s.t.: } \sum_{i=1}^{n}\alpha_i\, y^{(i)} = 0$$

$$0 \leq \alpha_i \leq C$$

At optimality, $w = \sum_i \alpha_i y^{(i)} x^{(i)}$, with

$$0 < \alpha_i < C \Longrightarrow y^{(i)}\left(w\cdot x^{(i)} + b\right) = 1$$

$$\alpha_i = C \Longrightarrow y^{(i)}\left(w\cdot x^{(i)} + b\right) = 1 - \xi_i$$