Duality

Sanjoy Dasgupta

University of California, San Diego

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 $\{(x^{(i)}, y^{(i)}) : i = 1 \ldots n\}$:

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 times an update occurred on point i.

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

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

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

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

- 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

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

$$\boxed{\begin{array}{lc} (\text{DUAL}) \quad \max_{\alpha \in \mathbb{R}^n} & \displaystyle\sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y^{(i)} y^{(j)} (x^{(i)} \cdot x^{(j)}) \\ & \text{s.t.:} \;\; \displaystyle\sum_{i=1}^n \alpha_i y^{(i)} = 0 \\ & \alpha \geq 0 \end{array}}$$

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

$$\alpha_i > 0 \quad \Rightarrow \quad y^{(i)}(w \cdot x^{(i)} + b) = 1$$

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

Dual of soft-margin SVM

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

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

$$\xi \geq 0$$

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

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

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

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

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

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