Formulating the optimization problem

The optimization problem

Problem: Given two linearly separable sets of points $A^{\pm} \subset \mathbb{R}^k$:

$$A^+ = \{x_1^+, \ldots, x_{n_+}\}$$

and

$$A^- = \{x_1^-, \ldots, x_{n_-}\}$$

Find points $p \in C(A^+)$ and $q \in C(A^-)$ so that

$$\|p - q\| = \min_{\substack{p' \in C(A^+) \\ q' \in C(A^-)}} \|p' - q'\|$$

The optimization problem (continued)

As above, let our linearly separable sets be

$$A^+ = \{x_1^+, \ldots, x_{n_+}\}$$

and

$$A^- = \{x_1^-, \ldots, x_{n_-}\}$$

Problem 1: Let $\lambda^\pm = (\lambda_1^\pm, \ldots, \lambda_{n_\pm}^\pm)$ be two vectors of real numbers of length $n_\pm$ respectively. Define

$$w(\lambda^+, \lambda^-) = \sum_{i=1}^{n_+} \lambda_i^+ x_i^+ - \sum_{i=1}^{n_-} \lambda_i^- x_i^- .$$

Find $\lambda^\pm$ such that $\|w(\lambda^+, \lambda^-)\|$ is minimal subject to the conditions that all $\lambda_i^\pm \geq 0$ and $\sum_{i=1}^{n_\pm} \lambda_i^\pm = 1$.

The optimization problem (continued)

Notice that:

- $w(\lambda^+, \lambda^-)$ is a quadratic function in the λ's with coefficients coming from the dot products of the $x_i^{\pm}$.

- The constraints are *inequalities* rather than equalities, so a direct application of Lagrange multipliers as we learned in calculus won't work.

- We will describe an iterative algorithm for solving this problem called Sequential Minimal Optimization, due to John Platt and introduced in 1998.

- First, though, we reformulate the problem slightly.

Reformulating the constrained optimization problem.

Problem 2: Let

$$Q(\lambda^+, \lambda^-) = \|w(\lambda^+, \lambda^-)\|^2 - \sum_{i=1}^{n_+} \lambda_i^+ - \sum_{i=1}^{n_-} \lambda_i^- .$$

Let $\lambda^\pm$ be values that minimize $Q(\lambda^+, \lambda^-)$ where all $\lambda_i^\pm \geq 0$ and

$$\alpha = \sum_{i=1}^{n_+} \lambda_i^+ = \sum_{i=1}^{n_-} \lambda_i^- > 0.$$

Then $\tau^\pm = (1/\alpha)\lambda^\pm$ is a solution to optimization problem 1.

Equivalence of the reformulated problem

Proof: We have (λ^+, λ^-) solving problem 2 and (σ^+, σ^-) solving problem 1; and finally we have $(\tau^+, \tau^-) = (1/\alpha)(\lambda^+, \lambda^-)$.

1. Since (τ^+, τ^-) satisfy the constraints of problem 2, we have:

$$Q(\lambda^+, \lambda^-) = \|w(\lambda^+, \lambda^-)\| - 2\alpha \leq \|w(\tau^+, \tau^-)\|^2 - 2$$

2. Since (τ^+, τ^-) satisfy the constraints of problem 1, we have

$$\|w(\sigma^+, \sigma^-)\|^2 \leq \|w(\tau^+, \tau^-)\|^2.$$

Equivalence of optimization problems continued

3. From (2), we have

$$\|w(\alpha\sigma^+, \alpha\sigma^-)\|^2 \leq \alpha^2\|w(\tau^+, \tau^-)\|^2 = \|w(\lambda^+, \lambda^-)\|^2$$

4. Subtracting 2α from both sides of this inequality yields

$$\|w(\alpha\sigma^+, \alpha\sigma^-)\|^2 - 2\alpha \leq Q(\lambda^+, \lambda^-).$$

Since (λ^+, λ^-) minimize $Q(\lambda^+, \lambda^-)$, this inequality must be an equality. Therefore

$$\alpha^2\|w(\sigma^+, \sigma^-)\| = \alpha^2\|w(\tau^+, \tau^-)\|^2$$

so (τ^+, τ^-) also gives a minimal value for problem 1.