Formulating the optimization problem

The optimization problem

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

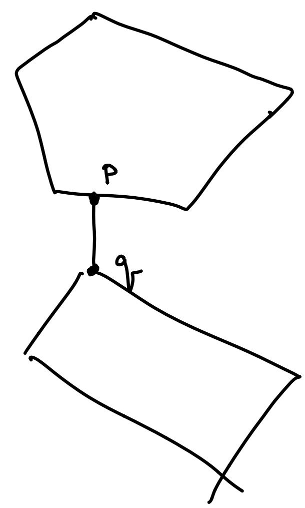
$$A^+ = \{x_1^+, \ldots, x_{n_+}^+\}$$

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

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

$\sum \lambda_i^+ x_i^+$ $\quad \sum \lambda_i^+ = 1$

$\lambda_i^+ \geq 0$

and

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

$\sum \lambda_i^- x_i^-$ $\quad \sum \lambda_i^- = 1$

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

$\lambda_i^- \geq 0$

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

Find $\lambda^\pm$ such that $\|w(\lambda^+, \lambda^-)\|^2$ 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 $x_i^\pm$.

$$w(\lambda^+,\lambda^-) = \left(\sum \lambda_i^+ x_i^+ - \sum \lambda_i^- x_i^-, \sum \lambda_i^+ x_i^+ - \sum \lambda_i^- x_i^-\right)$$

$$= \text{many terms } \lambda_i^\pm \lambda_j^\pm (x_i^\pm, x_j^\pm)$$

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

$$\lambda_i^+ \geqslant 0 \qquad 1 = \sum \lambda_i^+ = \sum \lambda_i^-$$

$$\lambda_i^- \geqslant 0$$

- 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.

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

Problem 2: Let

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

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

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

Then $\alpha \neq 0$ at the (λ^+, λ^-) that yield the minimum and $\tau^\pm = (1/\alpha)\lambda^\pm$ is a solution to optimization problem 1.

Equivalence of the reformulated problem

Proof:

$\omega(\lambda^+, \lambda^-) = \lambda x_1^+ - \lambda x_1^-$

First let all $\lambda_i^\pm = 0$ except $\lambda = \lambda_1^\pm$. Then

$$Q(\lambda^+, \lambda^-) = Q(\lambda) = \lambda^2 \|x_1^+ - x_1^-\|^2 - 2\lambda.$$

This takes its minimum value at $\lambda = 1/\|x_1^+ - x_1^-\|^2$ and at that point

$$Q(\lambda) = -\frac{1}{\|x_1^+ - x_1^-\|^2} < 0.$$

Therefore the minimum value is negative. But if $\alpha = 0$, then all $\lambda_i^\pm = 0$, so $Q = 0$ at such a point. Therefore $\alpha \neq 0$ at the minimum value.

To show the equivalence, 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:

$$Q(\lambda^+, \lambda^-) = \underbrace{\|w(\lambda^+, \lambda^-)\|^{2}}_{\text{minimum value}} - 2\alpha \leq \|w(\tau^+, \tau^-)\|^2 - 2$$

$(\tau^+, \tau^-) = \left(\frac{1}{\alpha}\lambda^+, \frac{1}{\alpha}\lambda^-\right)$

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

$$\underbrace{\|w(\sigma^+, \sigma^-)\|^2}_{\text{minimum value}} \leq \|w(\tau^+, \tau^-)\|^2.$$

Equivalence of optimization problems continued

3. From (2), we have

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

$$\|\sum \alpha \sigma_i^+ x_i^+ - \sum \alpha \sigma_i^- x_i^-\|^2 = \alpha^2 \|w(\sigma^+, \sigma^-)\|^2$$

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

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

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

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

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

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