Formulating the optimization problem

The optimization problem

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

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

and

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

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

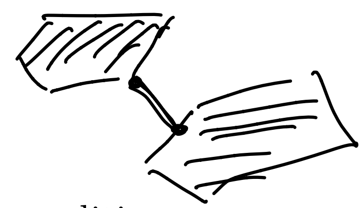
$$A^+ = \{x_1^+, \ldots, x_{n_+}^+\}$$

$$C(A^\pm) = \left\{ \sum_{i=1}^{n_\pm} \lambda_i^\pm x_i^\pm \;\middle|\; \sum \lambda_i^\pm = 1, \text{ all } \lambda_i^\pm \geq 0 \right\}.$$

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



$$\underline{w(\lambda^+, \lambda^-)} = \underbrace{\sum_{i=1}^{n_+} \lambda_i^+ x_i^+}_{\in C(A^+)} - \underbrace{\sum_{i=1}^{n_-} \lambda_i^- x_i^-}_{\in C(A^-)}.$$

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

The optimization problem (continued)

Notice that:

- $\|w(\lambda^+, \lambda^-)\|^2$ is a quadratic function in the λ's with coefficients coming from the dot products of the $x_i^{\pm}$.

$$\|w(\lambda^+,\lambda^-)\|^2 = \left(\sum \lambda_i^+ x_i^+ - \sum \lambda_i^- x_i^- , \sum \lambda_i^+ x_i^+ - \sum \lambda_i^- x_i^-\right)$$

typical term $\lambda_i^{\pm} \lambda_j^{\pm} (\underline{x_i^{\pm}, x_j^{\pm}})$

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

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

$$\lambda_i^{\pm} \geq 0.$$

- 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.

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 $(\tau^+, \tau^-) = (1/\alpha)(\lambda^+, \lambda^-)$.

minimize $\|w(\sigma^+,\sigma^-)\|^2$
$1 = \sum \sigma_i^+ = \sum \sigma_i^-$

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

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

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

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

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

$w(\sigma^+, \sigma^-) = \left(\sum \sigma_i^+ x_i^+ - \sum \sigma_i^- x_i^-, \sum \sigma_i^+ x_i^+ - \sum \sigma_i^- x_i^-\right)$

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

$\sum \sigma_i^+ = 1 \quad \sigma_i \geq 0$

$\sum \sigma_i^- = 1$

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

$$\longrightarrow \quad \underline{\|w(\alpha\sigma^+, \alpha\sigma^-)\|^2 - 2\alpha} \leq \underline{Q(\lambda^+, \lambda^-)}.$$ minimum value for Q

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

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

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

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