

Introduction to Machine Learning

Linear Support Vector Machines Support Vector Machine Training

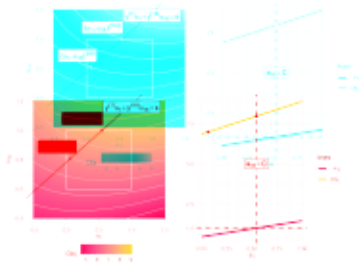


Learning goals

- Know that the SVM problem is differentiable

Learning goals

- Know that the SVM problem is not differentiable
- Know how to optimize the SVM problem in the primal via subgradient descent
- Know how to optimize SVM in the dual formulation via pairwise coordinate ascent
- Know how to optimize SVM in the dual formulation via pairwise coordinate ascent



TRAINING SVM IN THE DUAL / 2

Solution: Update two variables simultaneously

$$\begin{aligned} \max_{\alpha} \quad & \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y^{(i)} y^{(j)} \langle \mathbf{x}^{(i)}, \mathbf{x}^{(j)} \rangle \\ \text{s.t.} \quad & 0 \leq \alpha_i \leq C \quad \sum_{i=1}^n \alpha_i y^{(i)} = 0 \end{aligned}$$



Pairwise coordinate ascent in the dual

- 1: Initialize $\alpha = 0$ (or more cleverly)
 - 2: **for** $t = 1, 2, \dots$ **do**
 - 3: Select some pair α_i, α_j to update next
 - 4: Optimize dual w.r.t. α_i, α_j , while holding α_k ($k \neq i, j$) fixed
 - 5: **end for**
-

The objective is quadratic in the pair, and $s := y^{(i)}\alpha_i + y^{(j)}\alpha_j$ must stay constant. So both α are changed by same (absolute) amount, the signs of the change depend on the labels.

TRAINING SVM IN THE DUAL / 2

Sequential Minimal Optimization (SMO) exploits the fact that effectively we only need to solve a one-dimensional quadratic problem, over in interval, for which an analytical solution exists.

