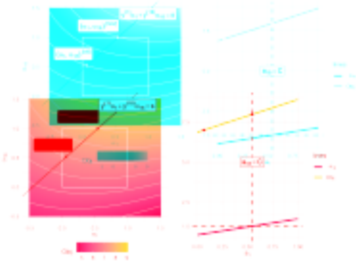
Introduction to Machine Learning

Linear Support Vector Machines

Support Vector Machine Training



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

TRAINING SVM IN THE DUAL / 2

Solution: Update two variables simultaneously

$$\max_{\alpha} \quad \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y^{(i)} y^{(j)} \left\langle \mathbf{x}^{(i)}, \mathbf{x}^{(j)} \right\rangle$$

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

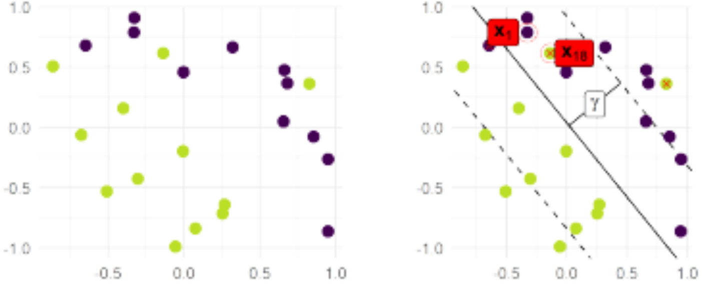
Pairwise coordinate ascent in the dual

```
Initialize α = 0 (or more cleverly)
for t = 1, 2, ... do
    Select some pair α_i, α_j to update next
    Optimize dual w.r.t. α_i, α_j, while holding α_k (k ≠ i, j) fixed
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 / 3

Assume we are in a valid state, $0 \leq \alpha_i \leq C$. Then we chose[1] two observations (encircled in red) for the next iteration. Note they have opposite labels so the sign of their change is equal.



[1] There are heuristics to pick the observations to speed up convergence.

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.