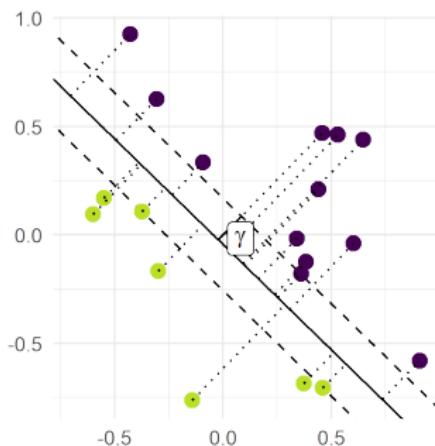


Introduction to Machine Learning

Linear Support Vector Machines Hard-Margin SVM Dual



Learning goals

- Know how to derive the SVM dual problem

HARD MARGIN SVM DUAL

We have derived the primal quadratic program for the hard margin SVM. We could directly solve this, but traditionally the SVM is solved in the dual and this has some advantages. In any case, many algorithms and derivations are based on it, so we need to know it.

$$\begin{aligned} \min_{\theta, \theta_0} \quad & \frac{1}{2} \|\theta\|^2 \\ \text{s.t.} \quad & y^{(i)} (\langle \theta, \mathbf{x}^{(i)} \rangle + \theta_0) \geq 1 \quad \forall i \in \{1, \dots, n\}. \end{aligned}$$

The Lagrange function of the SVM optimization problem is

$$\begin{aligned} L(\theta, \theta_0, \alpha) = \quad & \frac{1}{2} \|\theta\|^2 - \sum_{i=1}^n \alpha_i [y^{(i)} (\langle \theta, \mathbf{x}^{(i)} \rangle + \theta_0) - 1] \\ \text{s.t.} \quad & \alpha_i \geq 0 \quad \forall i \in \{1, \dots, n\}. \end{aligned}$$

The **dual** form of this problem is

$$\max_{\alpha} \min_{\theta, \theta_0} L(\theta, \theta_0, \alpha).$$



DUAL VARIABLE AND SUPPORT VECTORS

- SVs are defined to be points with $\hat{\alpha}_i > 0$. In the case of hard margin linear SVM, the SVs are on the edge of margin.
- However, not all points on edge of margin are necessarily SVs.
- In other words, it is possible that both $\hat{\alpha}_i = 0$ and $y^{(i)} (\langle \theta, \mathbf{x}^{(i)} \rangle) - 1 = 0$ hold.

