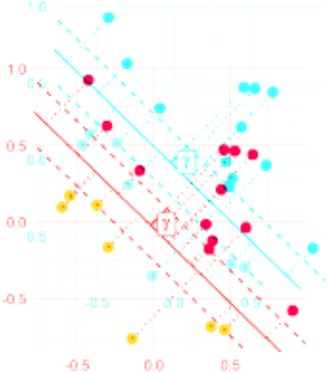
Introduction to Machine Learning

Linear Support Vector Machines

Hard-Margin SVM Dual

Hard-Margin SVM Dual



Learning goals

- Know how to derive the SVM dual problem

Learning goals

- Know how to derive the SVM dual problem

HARD MARGIN SVM DUAL / 2

Notice how the (p+1) decision variables $(\boldsymbol{\theta}, \theta_0)$ have become n decisions variables $\boldsymbol{\alpha}$, as constraints turned into variables and vice versa. Now every data point has an associated non-negative weight.

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

We find the stationary point of $L(\boldsymbol{\theta}, \theta_0, \boldsymbol{\alpha})$ w.r.t. $\boldsymbol{\theta}, \theta_0$ and obtain

$$
\begin{aligned}
\boldsymbol{\theta} &= \sum_{i=1}^{n} \alpha_i y^{(i)} \mathbf{x}^{(i)}, \\
0 &= \sum_{i=1}^{n} \alpha_i y^{(i)} \quad \forall\, i \in \{1, \dots, n\}.
\end{aligned}
$$

HARD MARGIN SVM DUAL / 3

By inserting these expressions & simplifying we obtain the dual problem

$$\begin{aligned} \max_{\boldsymbol{\alpha} \in \mathbb{R}^n} \quad & \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y^{(i)} y^{(j)} \left\langle \mathbf{x}^{(i)}, \mathbf{x}^{(j)} \right\rangle \\ \text{s.t.} \quad & \sum_{i=1}^{n} \alpha_i y^{(i)} = 0, \\ & \alpha_i \geq 0 \; \forall i \in \{1, \dots, n\}, \end{aligned}$$

or, equivalently, in matrix notation:

$$\begin{aligned} \max_{\boldsymbol{\alpha} \in \mathbb{R}^n} \quad & \mathbf{1}^T \boldsymbol{\alpha} - \frac{1}{2} \boldsymbol{\alpha}^T \operatorname{diag}(\mathbf{y}) \boldsymbol{K} \operatorname{diag}(\mathbf{y}) \boldsymbol{\alpha} \\ \text{s.t.} \quad & \boldsymbol{\alpha}^T \mathbf{y} = 0, \\ & \boldsymbol{\alpha} \geq 0, \end{aligned}$$

with $\boldsymbol{K} := \mathbf{X}\mathbf{X}^T$.

HARD MARGIN SVM DUAL / 4

If $(\boldsymbol{\theta}, \theta_0, \boldsymbol{\alpha})$ fulfills the KKT conditions (stationarity, primal/dual feasibility, complementary slackness), it solves both the primal and dual problem (strong duality).
Under these conditions, and if we solve the dual problem and obtain $\hat{\boldsymbol{\alpha}}$, we know that $\boldsymbol{\theta}$ is a linear combination of our data points:

$$\hat{\boldsymbol{\theta}} = \sum_{i=1}^{n} \hat{\alpha}_i y^{(i)} \mathbf{x}^{(i)}$$

Complementary slackness means:

$$\hat{\alpha}_i \left[y^{(i)} \left(\left\langle \boldsymbol{\theta}, \mathbf{x}^{(i)} \right\rangle + \theta_0 \right) - 1 \right] = 0 \quad \forall\, i \in \{1, ..., n\}.$$

HARD MARGIN SVM DUAL / 5

$$\hat{\boldsymbol{\theta}} = \sum_{i=1}^{n} \hat{\alpha}_i y^{(i)} \mathbf{x}^{(i)}$$

$$\hat{\alpha}_i \left[y^{(i)} \left(\left\langle \boldsymbol{\theta}, \mathbf{x}^{(i)} \right\rangle + \theta_0 \right) - 1 \right] = 0 \quad \forall\, i \in \{1, ..., n\}.$$

- So either $\hat{\alpha}_i = 0$, and is not active in the linear combination, or $\hat{\alpha}_i > 0$, then $y^{(i)} \left(\langle \boldsymbol{\theta}, \mathbf{x}^{(i)} \rangle + \theta_0 \right) = 1$, and $(\mathbf{x}^{(i)}, y^{(i)})$ has minimal margin and is a support vector!
- We see that we can directly extract the support vectors from the dual variables and the $\boldsymbol{\theta}$ solution only depends on them.
- We can reconstruct the bias term θ_0 from any support vector:

$$\theta_0 = y^{(i)} - \left\langle \boldsymbol{\theta}, \mathbf{x}^{(i)} \right\rangle.$$