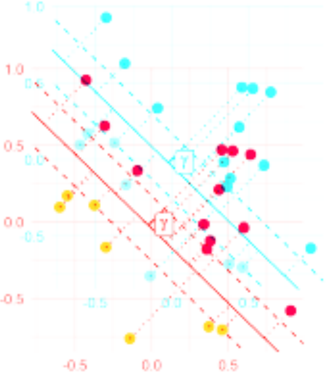
Introduction to Machine Learning

Linear Support Vector Machines

Linear Hard Margin SVM



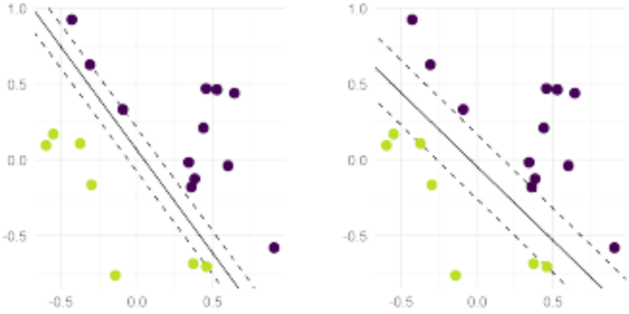
Learning goals

- Know that the hard-margin SVM maximizes the margin between data points and hyperplane
- Know that this is a quadratic program
- Know that support vectors are the data points closest to the separating hyperplane

Learning goals

- Know that the hard-margin SVM maximizes the margin between data points and hyperplane
- Know that this is a quadratic program
- Know that support vectors are the data points closest to the separating hyperplane

LINEAR CLASSIFIERS / 2



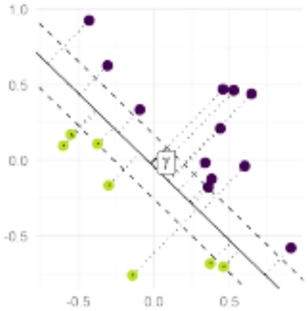
- We want study how to build a binary, linear classifier from solid geometrical principles.
- Which of these two classifiers is “better”?

$\rightarrow$ The decision boundary on the right has a larger **safety margin.**

SUPPORT VECTOR MACHINES: GEOMETRY / 2

- $$d\left(f, \mathbf{x}^{(i)}\right) = \frac{y^{(i)} f\left(\mathbf{x}^{(i)}\right)}{\|\boldsymbol{\theta}\|} = y^{(i)} \frac{\boldsymbol{\theta}^T \mathbf{x}^{(i)} + \theta_0}{\|\boldsymbol{\theta}\|}$$

  computes the (signed) distance to the separating hyperplane $f(\mathbf{x}) = 0$, positive for correct classifications, negative for incorrect.
- This expression becomes negative for misclassified points.

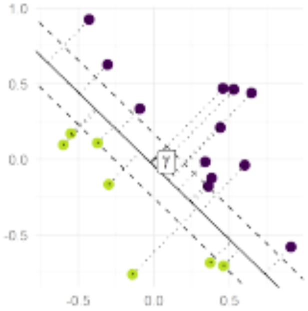


SUPPORT VECTOR MACHINES: GEOMETRY / 3

- The distance of f to the whole dataset $\mathcal{D}$ is the smallest distance

$$\gamma = \min_i \left\{ d\left(f, \mathbf{x}^{(i)}\right) \right\}.$$

- This represents the “safety margin”, it is positive if f separates and we want to maximize it.



MAXIMUM MARGIN SEPARATION / 2

- Note that the same hyperplane does not have a unique representation:

$$\{\mathbf{x} \in \mathcal{X} \mid \boldsymbol{\theta}^\top \mathbf{x} = 0\} = \{\mathbf{x} \in \mathcal{X} \mid c \cdot \boldsymbol{\theta}^\top \mathbf{x} = 0\}$$

  for arbitrary $c \neq 0$.
- To ensure uniqueness of the solution, we make a reference choice – we only consider hyperplanes with $\|\boldsymbol{\theta}\| = 1/\gamma$:

$$\begin{aligned} \max_{\boldsymbol{\theta}, \theta_0} \quad & \gamma \\ \text{s.t.} \quad & y^{(i)} \left(\left\langle \boldsymbol{\theta}, \mathbf{x}^{(i)} \right\rangle + \theta_0 \right) \geq 1 \quad \forall i \in \{1, \dots, n\}. \end{aligned}$$

MAXIMUM MARGIN SEPARATION / 3

- Substituting $\gamma = 1/\|\boldsymbol{\theta}\|$ in the objective yields:

$$\max_{\boldsymbol{\theta},\theta_0} \quad \frac{1}{\|\boldsymbol{\theta}\|}$$
$$\text{s.t.} \quad y^{(i)}\left(\left\langle \boldsymbol{\theta}, \mathbf{x}^{(i)}\right\rangle + \theta_0\right) \geq 1 \quad \forall i \in \{1, \dots, n\}.$$

- Maximizing $1/\|\boldsymbol{\theta}\|$ is the same as minimizing $\|\boldsymbol{\theta}\|$, which is the same as minimizing $\frac{1}{2}\|\boldsymbol{\theta}\|^2$:

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