Support vector machines

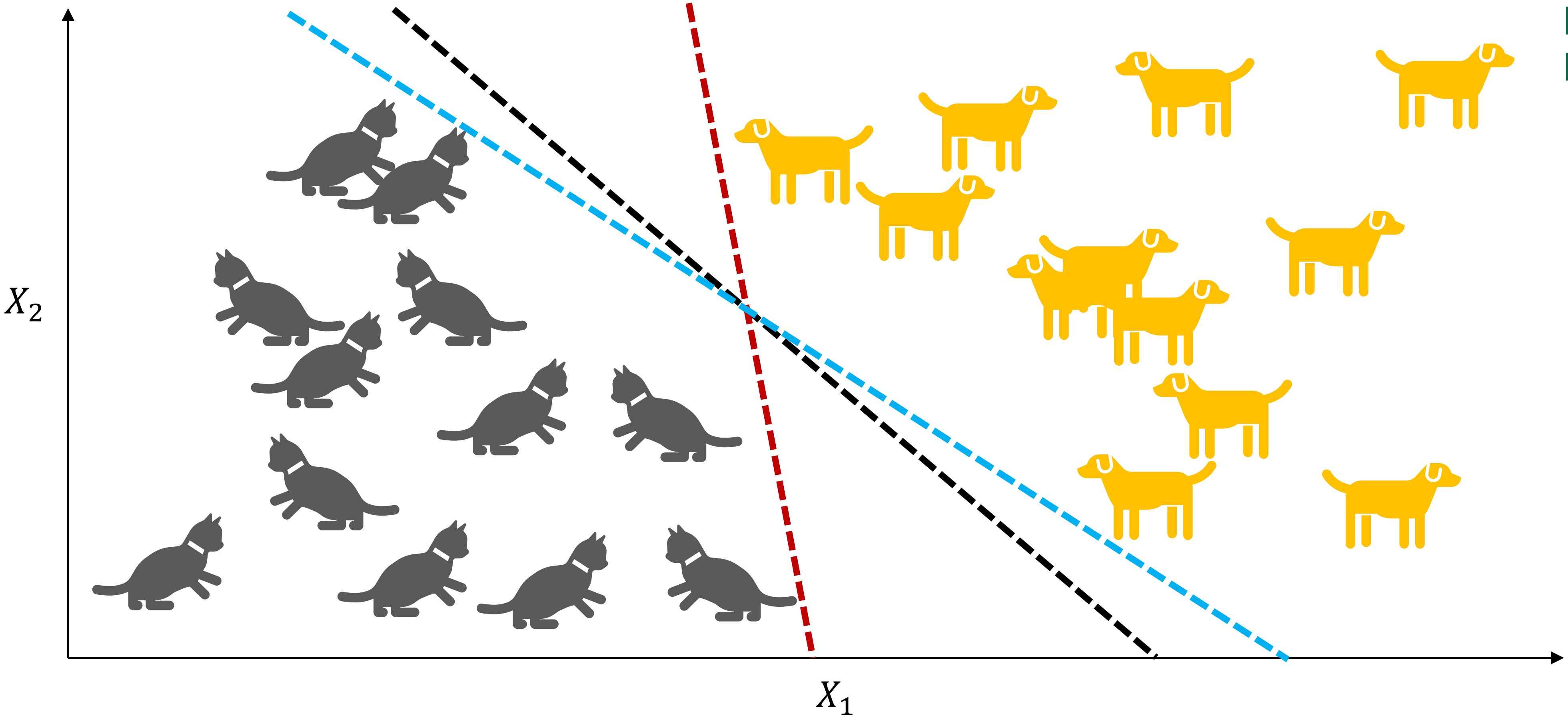
Lecture 12 of “Mathematics and AI”

Outline

1. Maximal margin classifier
2. Support vector classifier
3. SVC and logistic regression

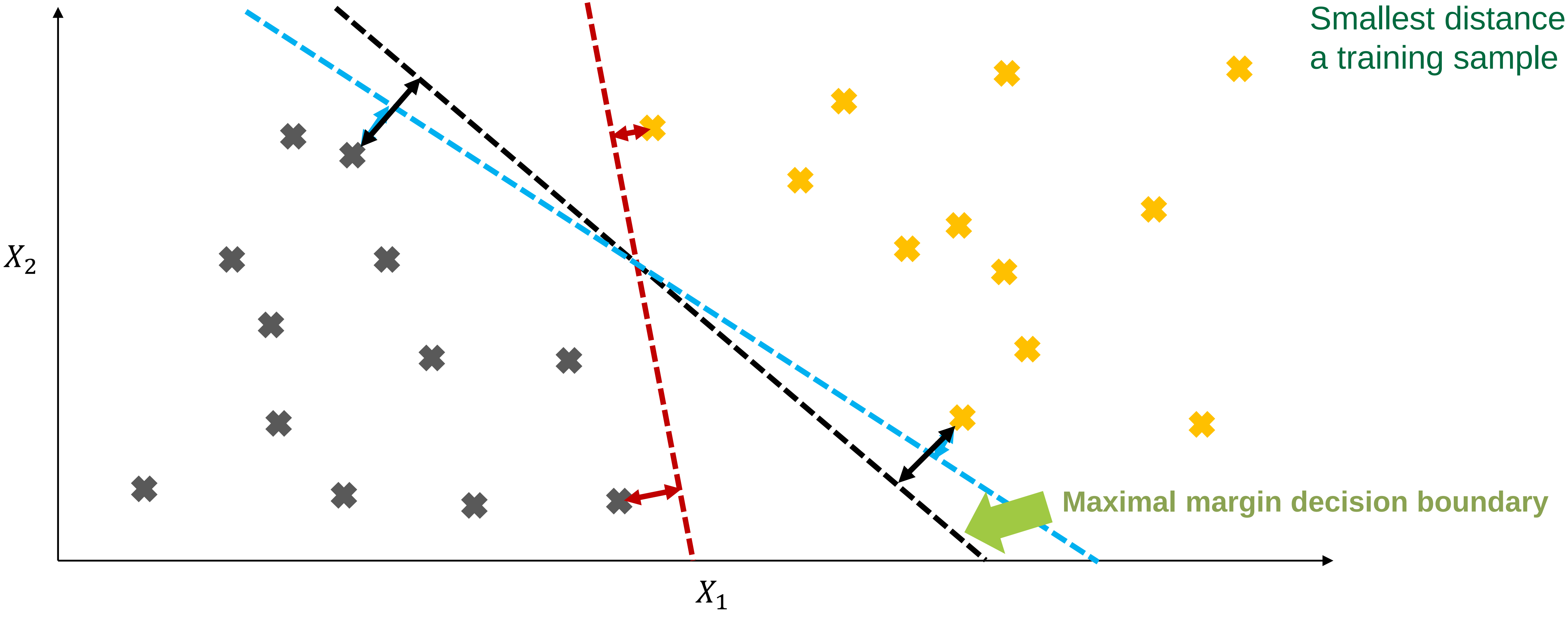
Maximal margin classifier

Linearly separable patterns



Which decision boundary is the best?

Linearly separable patterns



Margin:
Smallest distance to a training sample

Hyperplanes

- Hyperplane (perpendicular to coordinate axis)

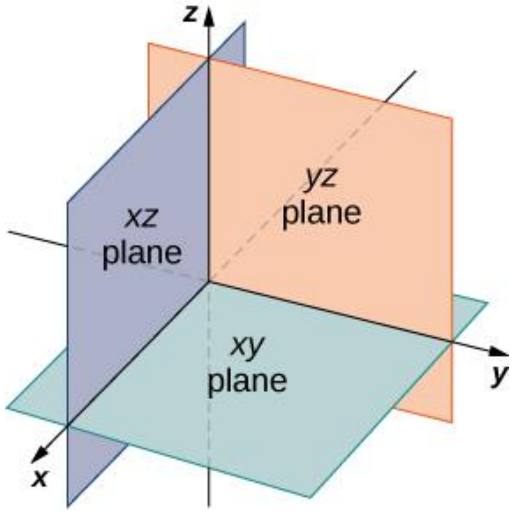
$$\beta_0 + x_j = 0$$

- Hyperplane (general case)

$$\beta_0 + \beta_1 x_{i,1} + \beta_2 x_{i,2} + \cdots + \beta_p x_{i,p} = 0$$

- Hyperplane in vector notation

$$\beta_0 + \vec{\beta} \cdot \vec{x_i} = 0$$



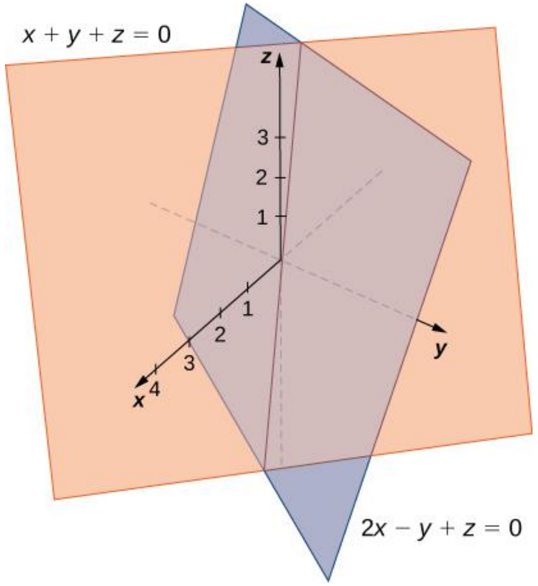


Image source: OpenStax Calculus 3

Hyperplanes as separators

- Hyperplane in vector notation

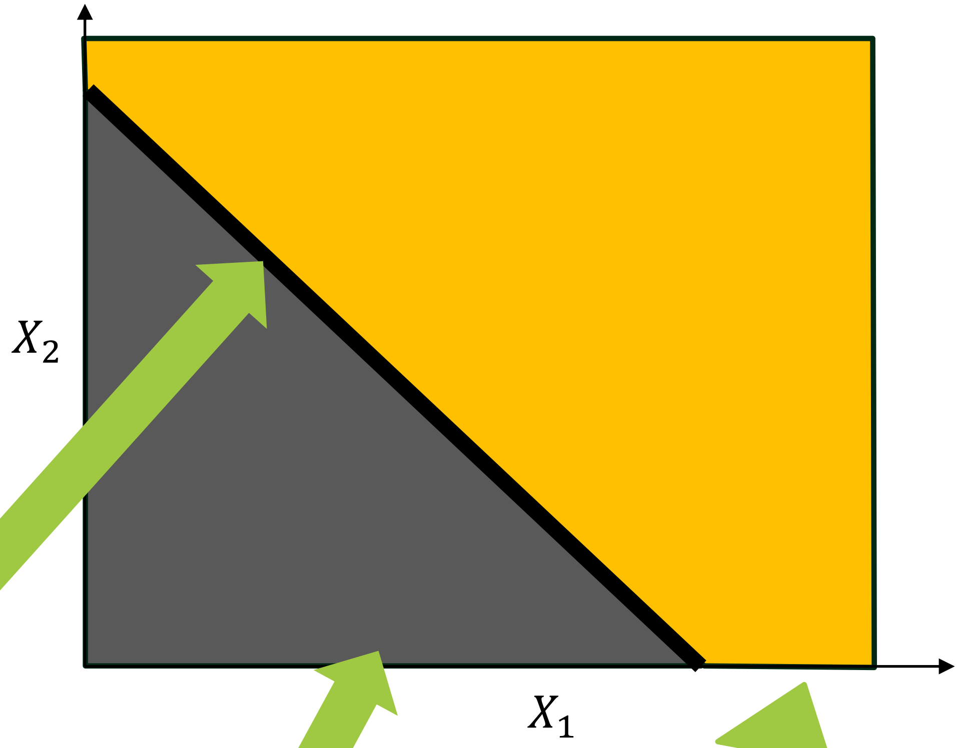
$$\beta_0 + \vec{\beta} \cdot \vec{x_i} = 0$$

- Three sets of points

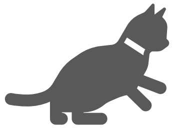
$\left\{ \vec{x_i} \mid \beta_0 + \vec{\beta} \cdot \vec{x_i} = 0 \right\} \rightarrow \vec{x_i}$ in hyperplane,

$\left\{ \vec{x_i} \mid \beta_0 + \vec{\beta} \cdot \vec{x_i} < 0 \right\} \rightarrow \vec{x_i}$ on one side of hyperplane,

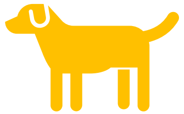
$\left\{ \vec{x_i} \mid \beta_0 + \vec{\beta} \cdot \vec{x_i} > 0 \right\} \rightarrow \vec{x_i}$ on other side of hyperplane



Linear decision boundary for binary classification

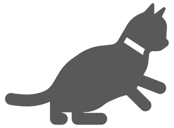
$\left\{ \overrightarrow{x_i} \mid \beta_0 + \vec{\beta} \cdot \overrightarrow{x_i} < 0 \right\}$ 

- $\overrightarrow{x_i}$ on one side of hyperplane
- $(\overrightarrow{x_i}, y_i)$ is "cat"
- $y_i = -1$

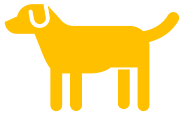
$\left\{ \overrightarrow{x_i} \mid \beta_0 + \vec{\beta} \cdot \overrightarrow{x_i} > 0 \right\}$ 

- $\overrightarrow{x_i}$ on other side of hyperplane
- $(\overrightarrow{x_i}, y_i)$ is "dog"
- $y_i = +1$

For **<u>all</u>** training samples:



$$y_i \left(\beta_0 + \vec{\beta} \cdot \overrightarrow{x_i} \right) > 0$$



Replace with quantitative formulation?

Hyperplanes and normal vectors

- Hyperplane in vector notation

$$\beta_0 + \vec{\beta} \cdot \overrightarrow{x_i} = 0$$

- Hyperplane in normal-vector notation

$$\vec{\beta} \cdot \left(\overrightarrow{x_i} - \vec{P}\right) = 0$$

$$\text{with } \vec{P} = \begin{pmatrix} -\frac{\beta_0}{\beta_1} & 0 & \ldots & 0 \end{pmatrix}^T$$

- Vector $\vec{\beta}$ is the **normal vector** of the hyperplane

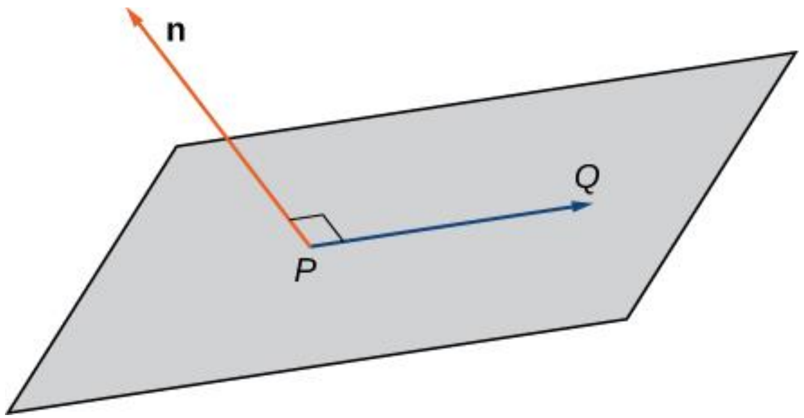


Image source: OpenStax Calculus 3

Normal vectors and margins

- Hyperplane in normal-vector notation

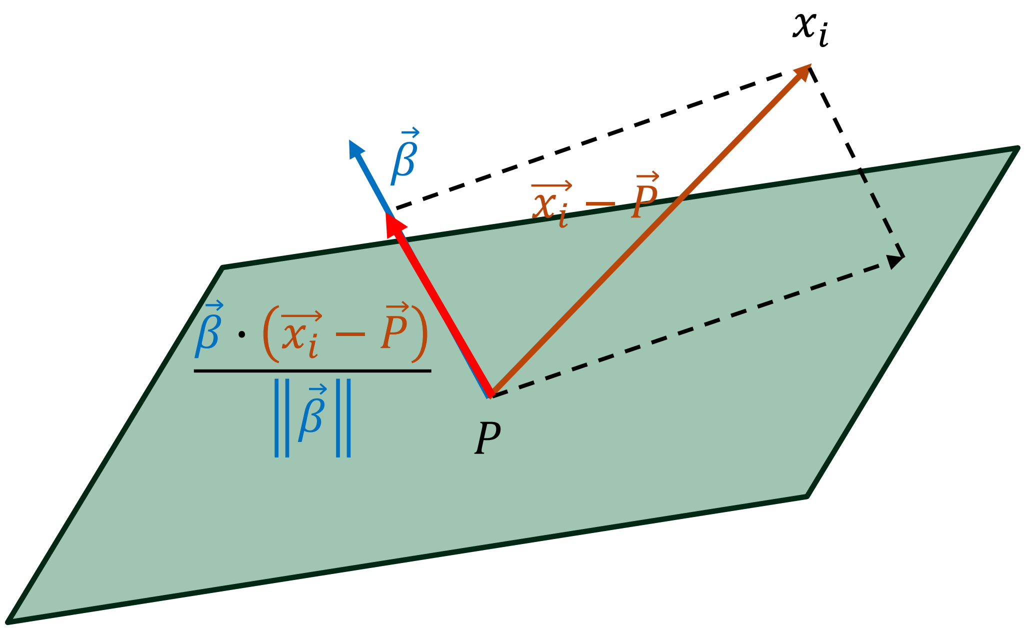
$$\vec{\beta} \cdot \left(\vec{x_i} - \vec{P}\right) = 0$$

- Distance from $\vec{x_i}$ to hyperplane

$$d_{\vec{\beta},\vec{x_i}} = \frac{\left\|\vec{\beta} \cdot \left(\vec{x_i} - \vec{P}\right)\right\|}{\left\|\vec{\beta}\right\|}$$

- Special case when $\left\|\vec{\beta}\right\| = 1$:

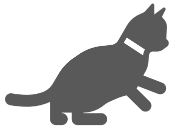
$$d_{\vec{\beta},\vec{x_i}} = \left\|\vec{\beta} \cdot \left(\vec{x_i} - \vec{P}\right)\right\| = \left|\beta_0 + \vec{\beta} \cdot \vec{x_i}\right|$$



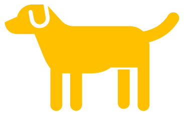
Maximal margin classifier

- Find the linear decision boundary (i.e., hyperplane) with the largest margin

For **all** training samples:



$$y_i \left(\beta_0 + \vec{\beta} \cdot \overrightarrow{x_i}\right) > M$$



$$\text{with } \left\|\vec{\beta}\right\| = 1$$

Maximal margin classifier: Optimization problem

$$\underset{\beta_0,\beta_1,\ldots,\beta_p,M}{\text{maximize}} \; M$$

$$\text{subject to } \sum_{j=1}^{p} \beta_j^2 = 1,$$

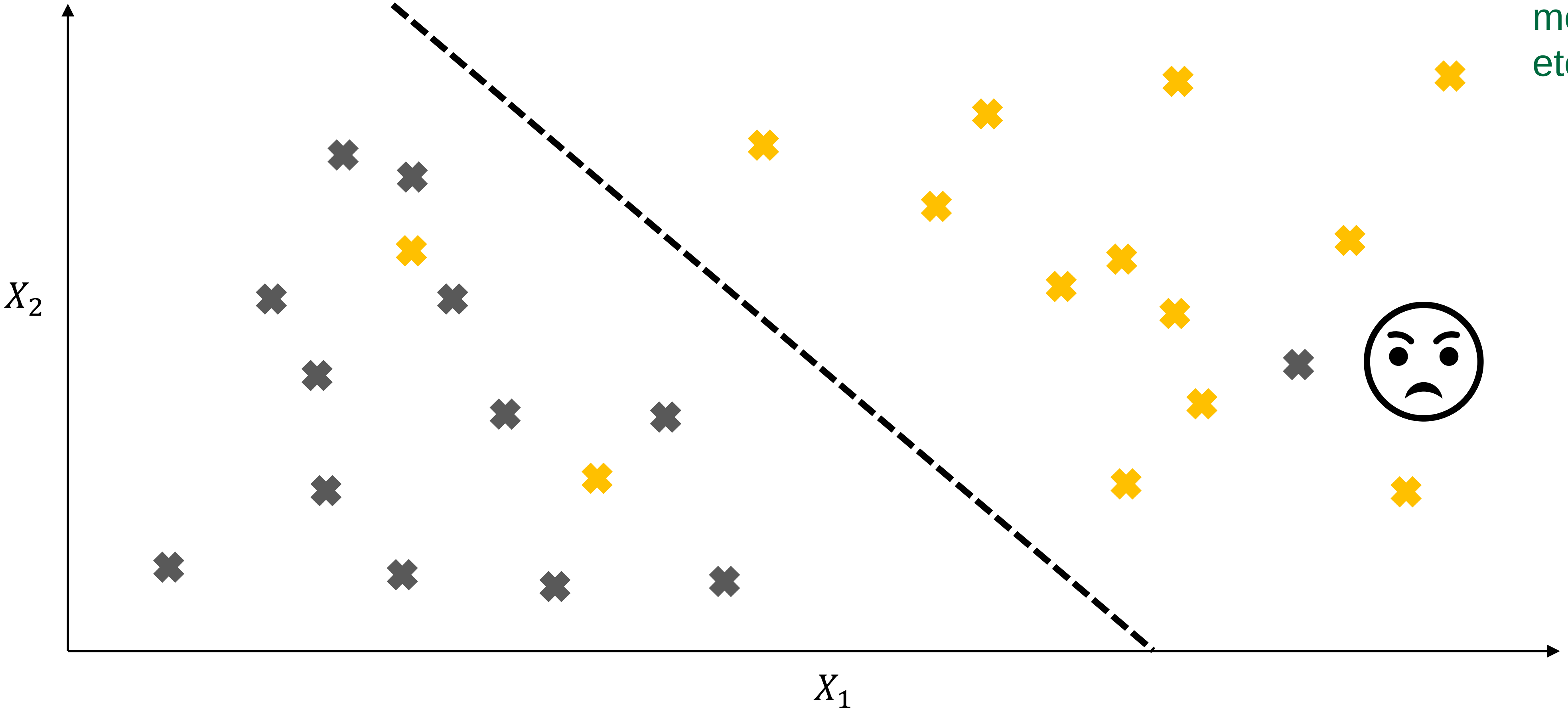
$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}) \geq M \;\; \forall\, i = 1, \ldots, n$$

Image source: Introduction to Statistical Learning in Python

Alternative optimization problem

Support vector classifier

Data without a linear separator



Intrinsic variability, measurement errors, etc.

Slack variables ε_i

$$\underset{\beta_0,\beta_1,\ldots,\beta_p,\epsilon_1,\ldots,\epsilon_n,\, M}{\text{maximize}} \quad M$$

$$\text{subject to} \quad \sum_{j=1}^{p} \beta_j^2 = 1,$$

$$y_i(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}) \geq M \boxed{(1 - \epsilon_i)},$$

$$\boxed{\epsilon_i \geq 0, \quad \sum_{i=1}^{n} \epsilon_i \leq C,}$$

Slack variable

Tuning parameter / hyperparameter

Image source: Introduction to Statistical Learning in Python

DARTMOUTH

Examples

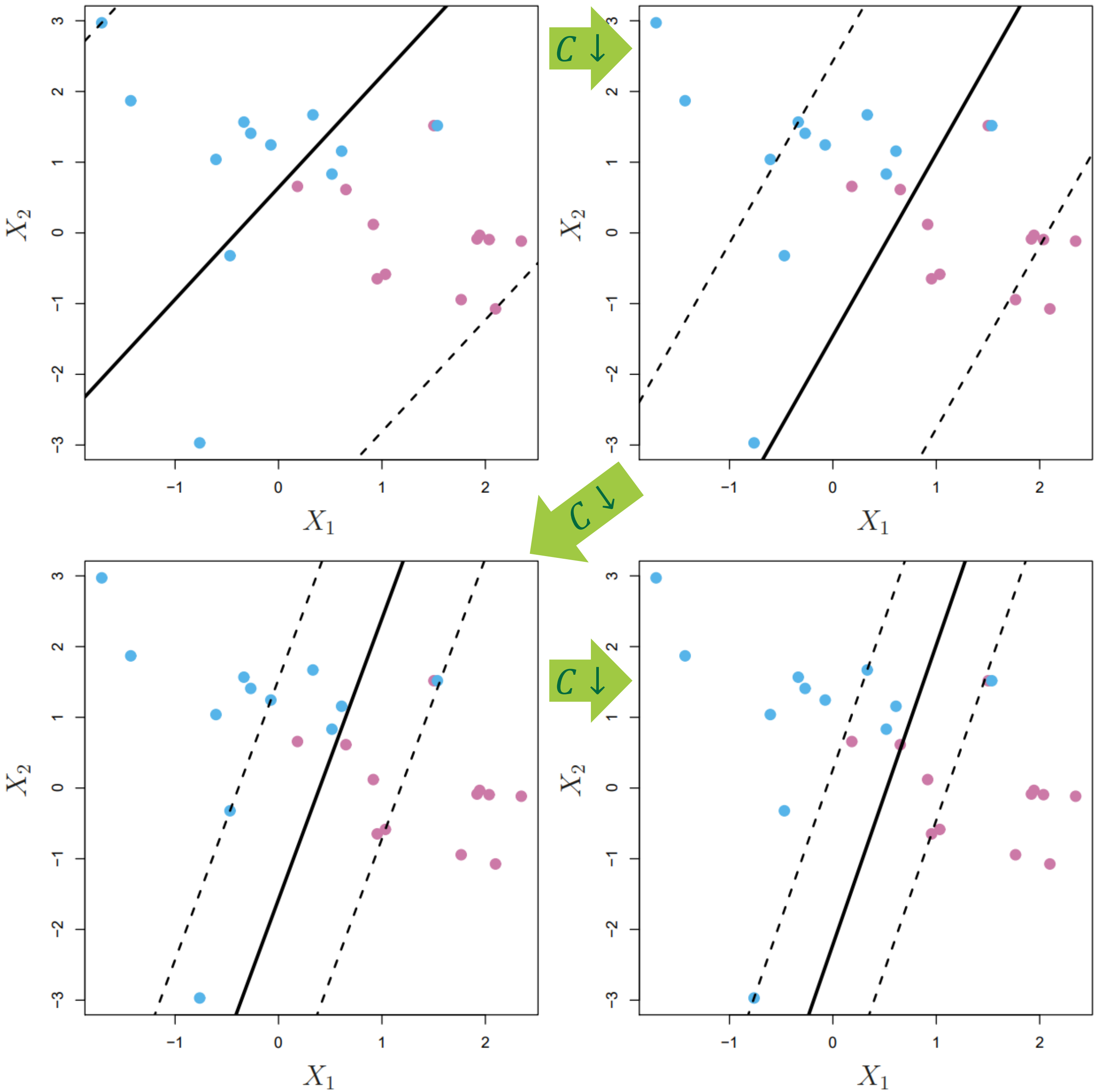


Image source: Introduction to Statistical Learning in Python

SVC & logistic regression

Alternative optimization problem for SVC

Loss functions for SVC and linear regression