Statistical Learning STA4042

DECISION BOUNDARIES

SCHOOL OF DATA SCIENCE

Linear and non-linear decision

Concerns Unsupervised and Supervised learning

Although more pictural with Classification also concerns Regression

Statistical Learning STA4042

DECISION BOUNDARIES

SCHOOL OF DATA SCIENCE

Linear and non-linear decision

Concerns Unsupervised and Supervised learning

Although more pictural with Classification also concerns Regression

Examples of linear decision:

- Linear Regression (Least-square, Poisson, logistic, softmax..)
- Naive Bayes, Linear discriminent analysis,
- Methods from PCA if only one principal component used
- Simple support vector machine

Statistical Learning STA4042

DECISION BOUNDARIES

SCHOOL OF DATA SCIENCE

Linear and non-linear decision

Concerns Unsupervised and Supervised learning

Although more pictural with Classification also concerns Regression

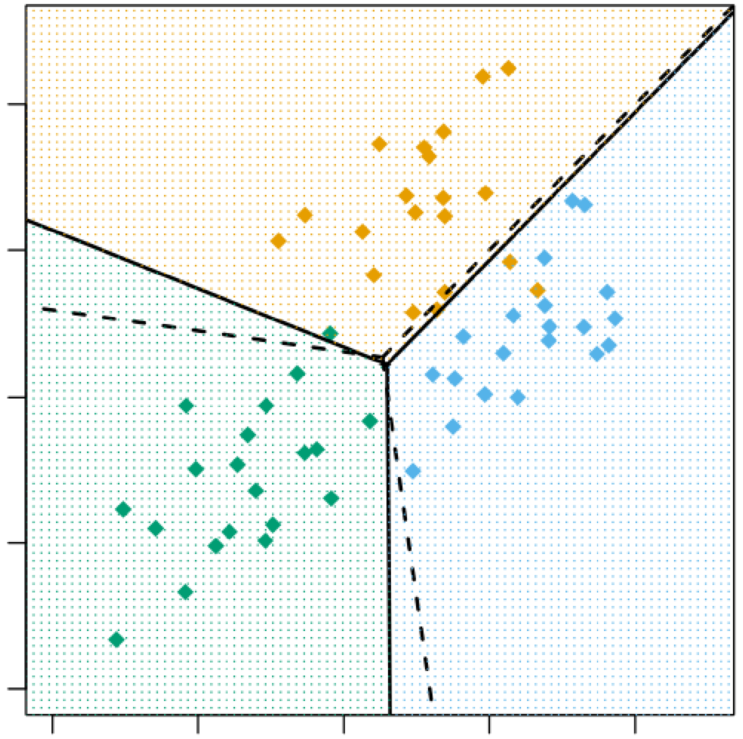
Examples of linear decision:

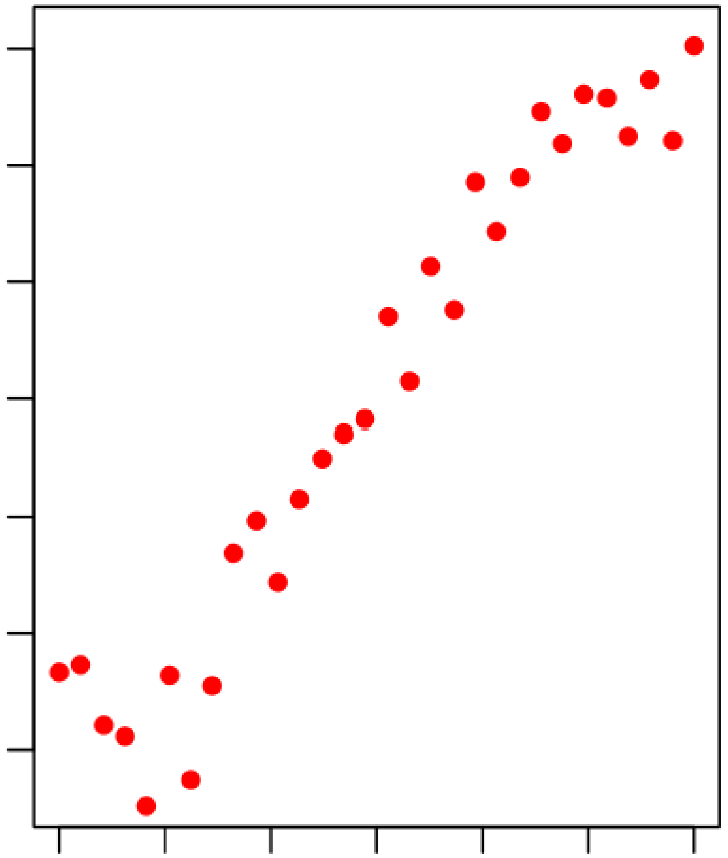
- Linear Regression (Least-square, Poisson, logistic, softmax..)
- Naive Bayes, Linear discriminent analysis,
- Methods from PCA if only one principal component used
- Simple support vector machine

Examples of non-linear decisions:

- K nearest neighbour
- QDA, Polynomial fits.
- k-means
- Kernel methods (spectral clustering, SVM)
- Neural networks

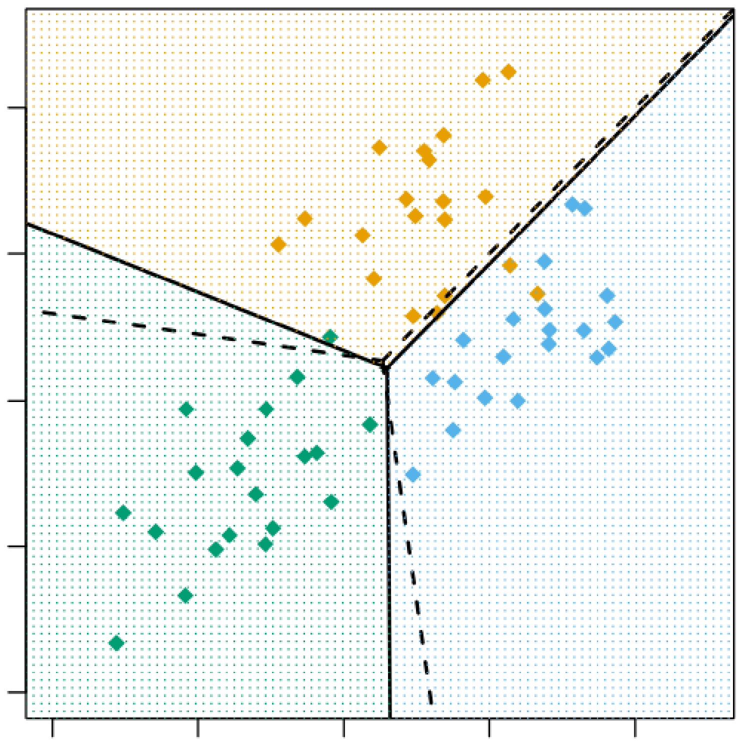
Linear and non-linear decision

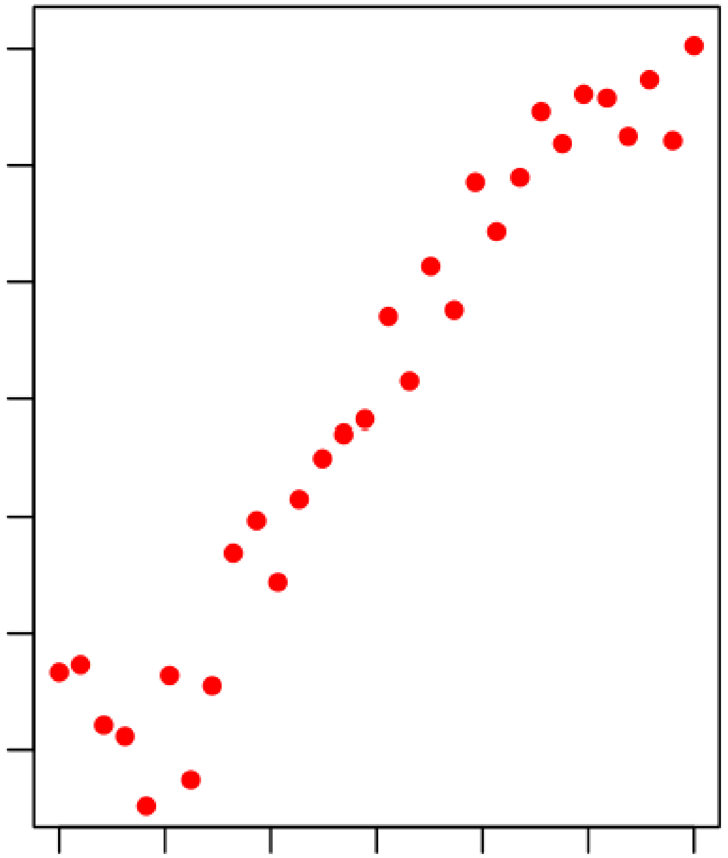




Linear Decisions

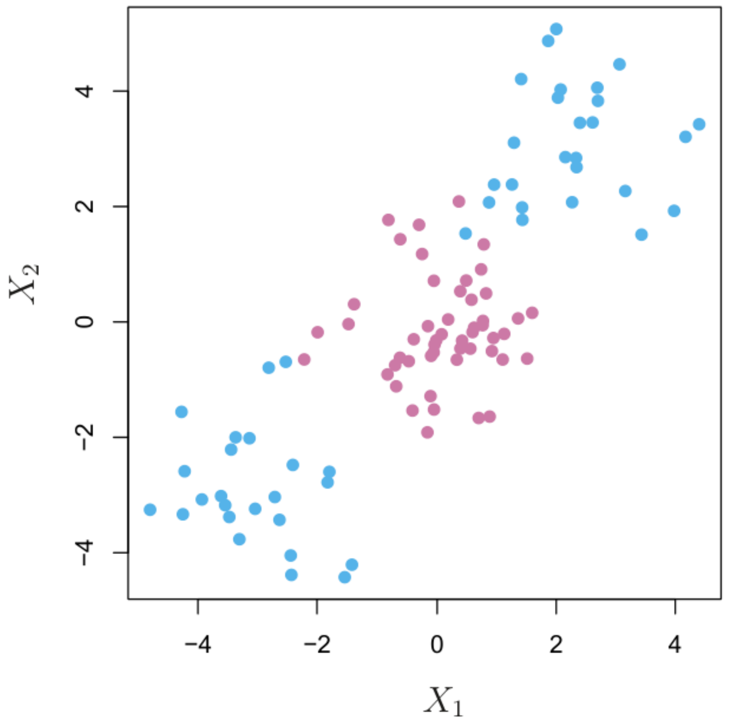
Linear and non-linear decision

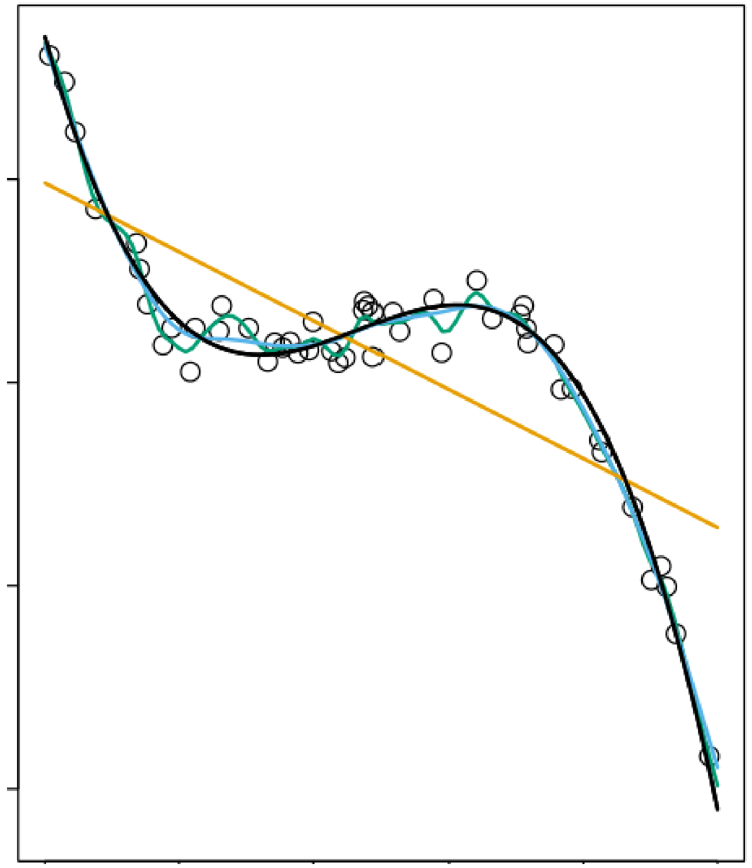




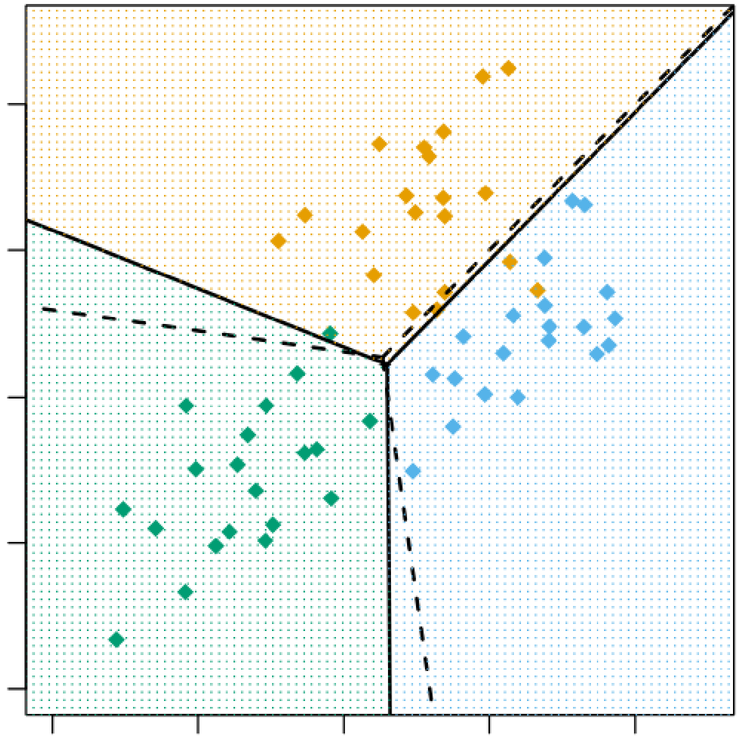
Linear Decisions

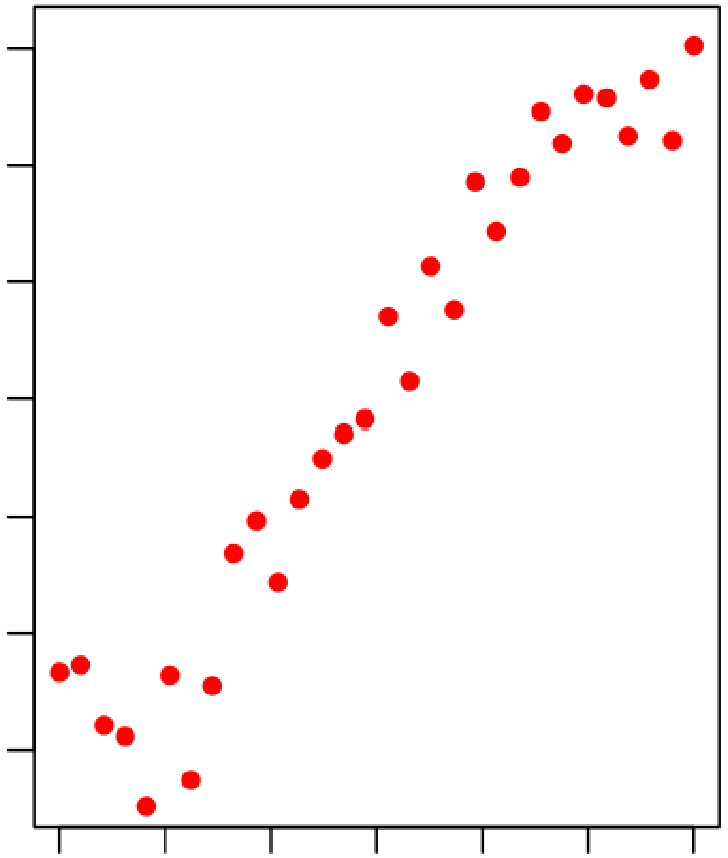
Non-linear decisions





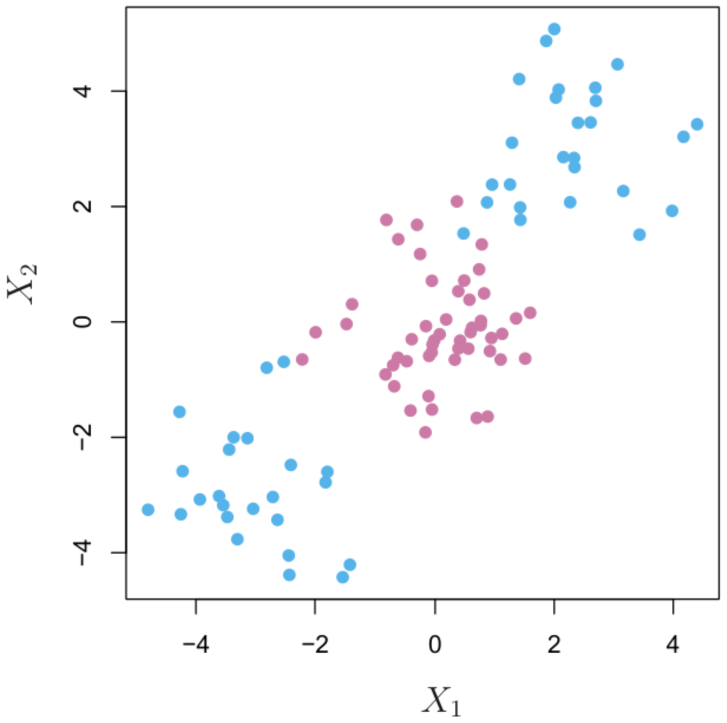
Linear and non-linear decision

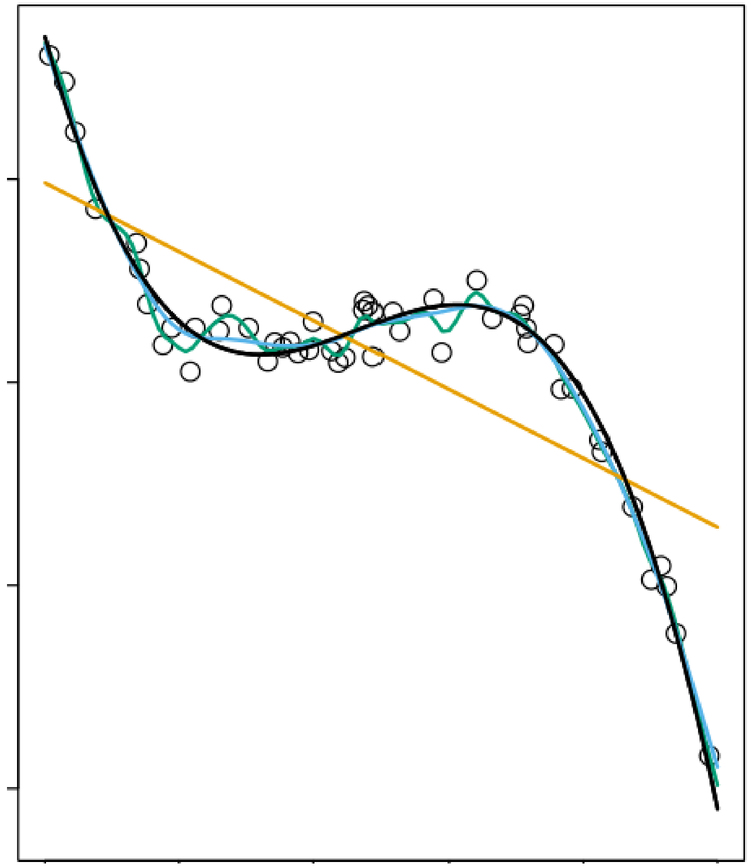




Linear Decisions

Non-linear decisions

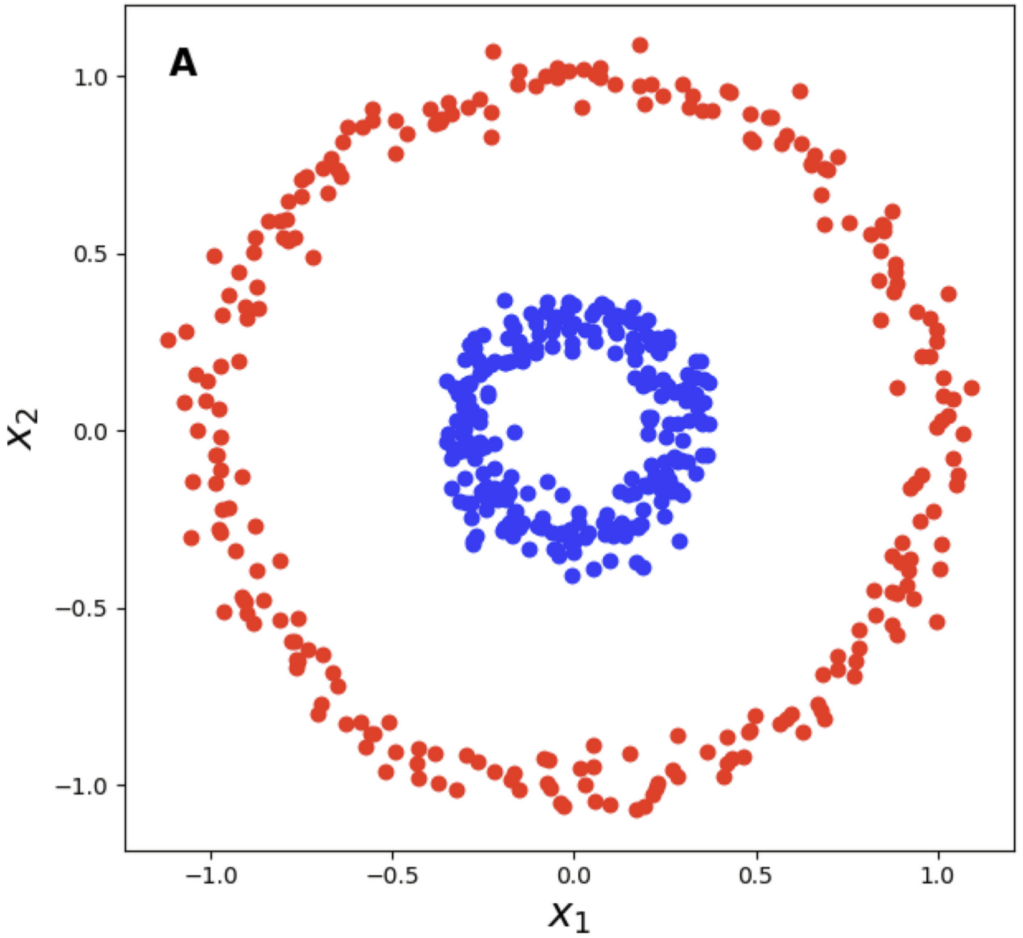




Question even more important with bigger dimension

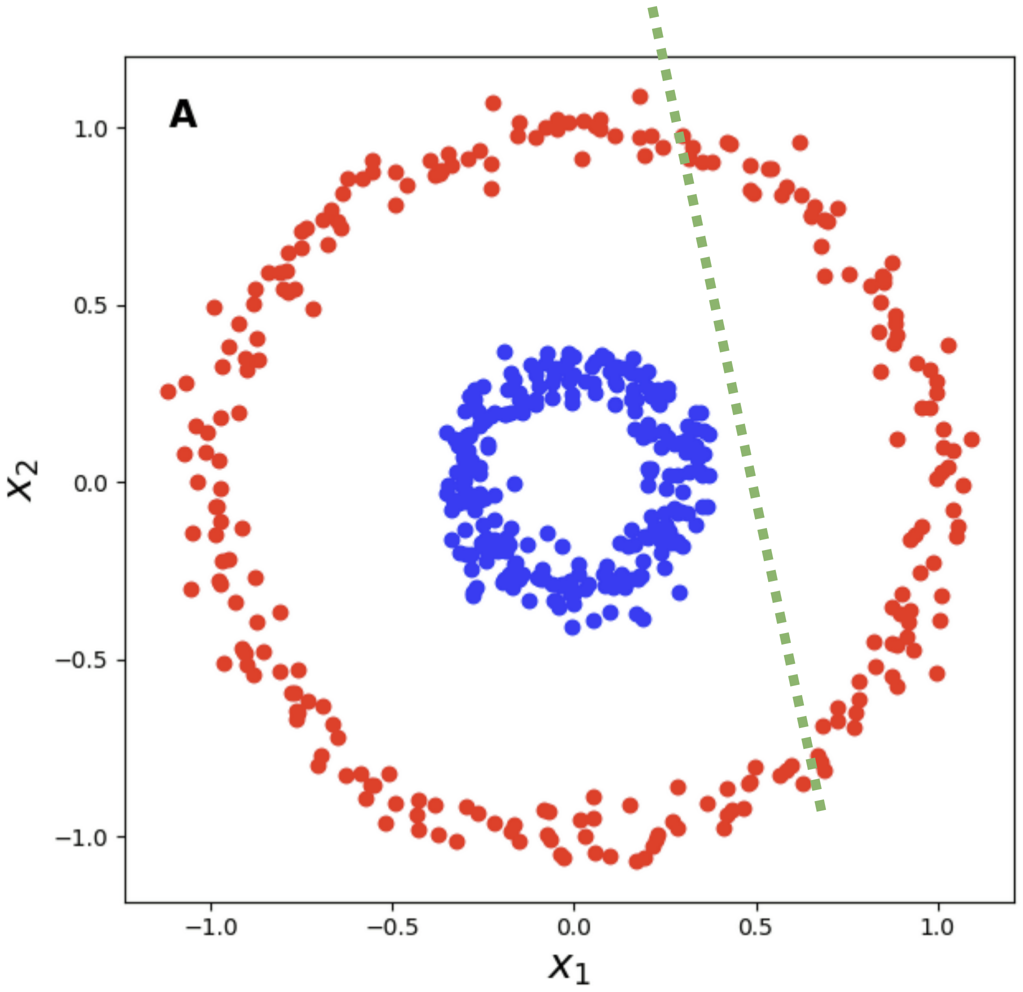
How to deal with non linear boundaries ? *(Classification or regression)*

Linear method can not provide good solution



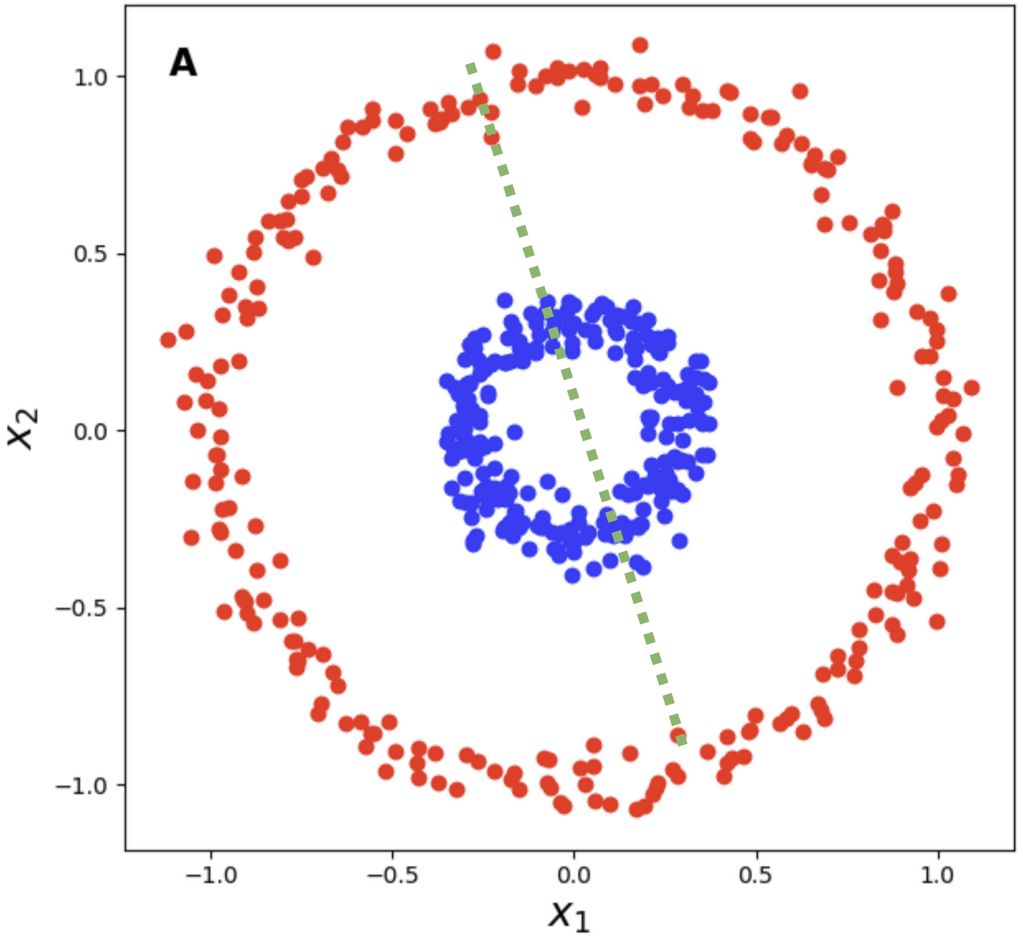
How to deal with non linear boundaries ? *(Classification or regression)*

Linear method can not provide good solution



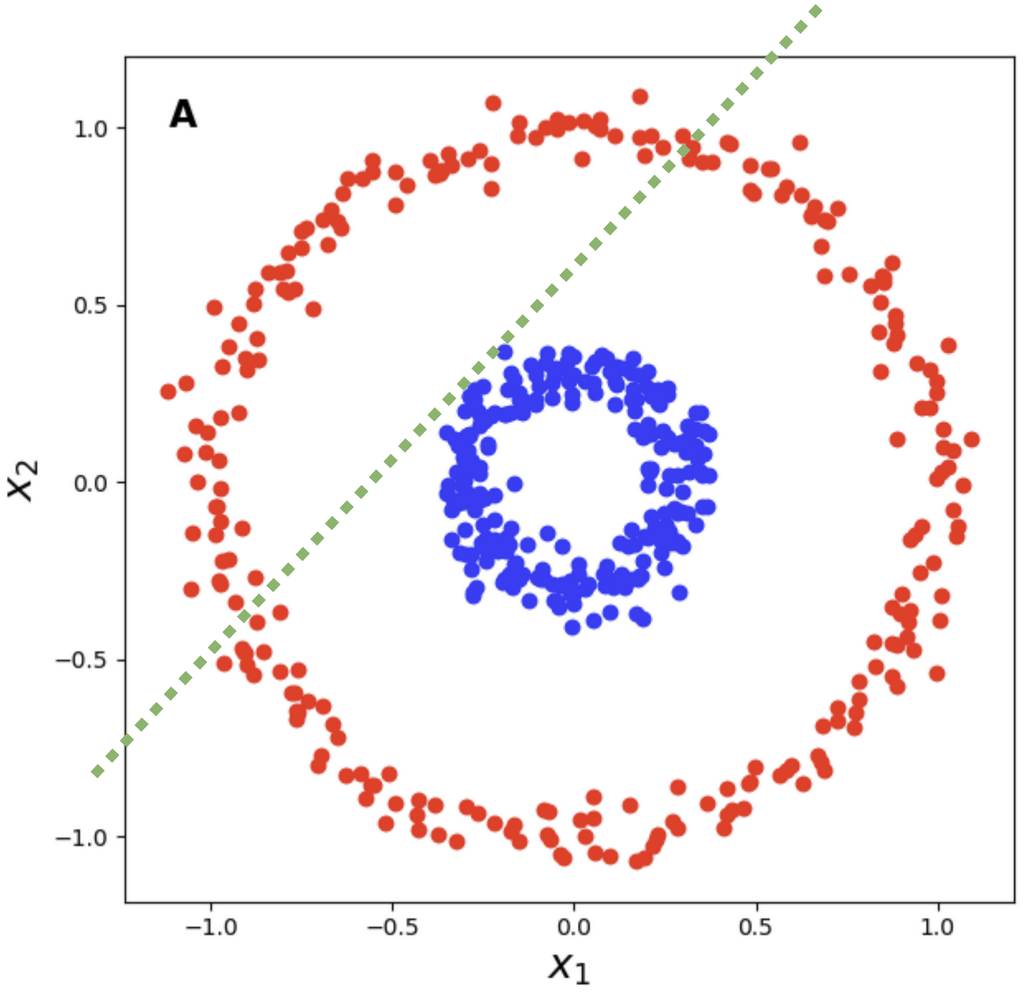
How to deal with non linear boundaries ? *(Classification or regression)*

Linear method can not provide good solution



How to deal with non linear boundaries ? *(Classification or regression)*

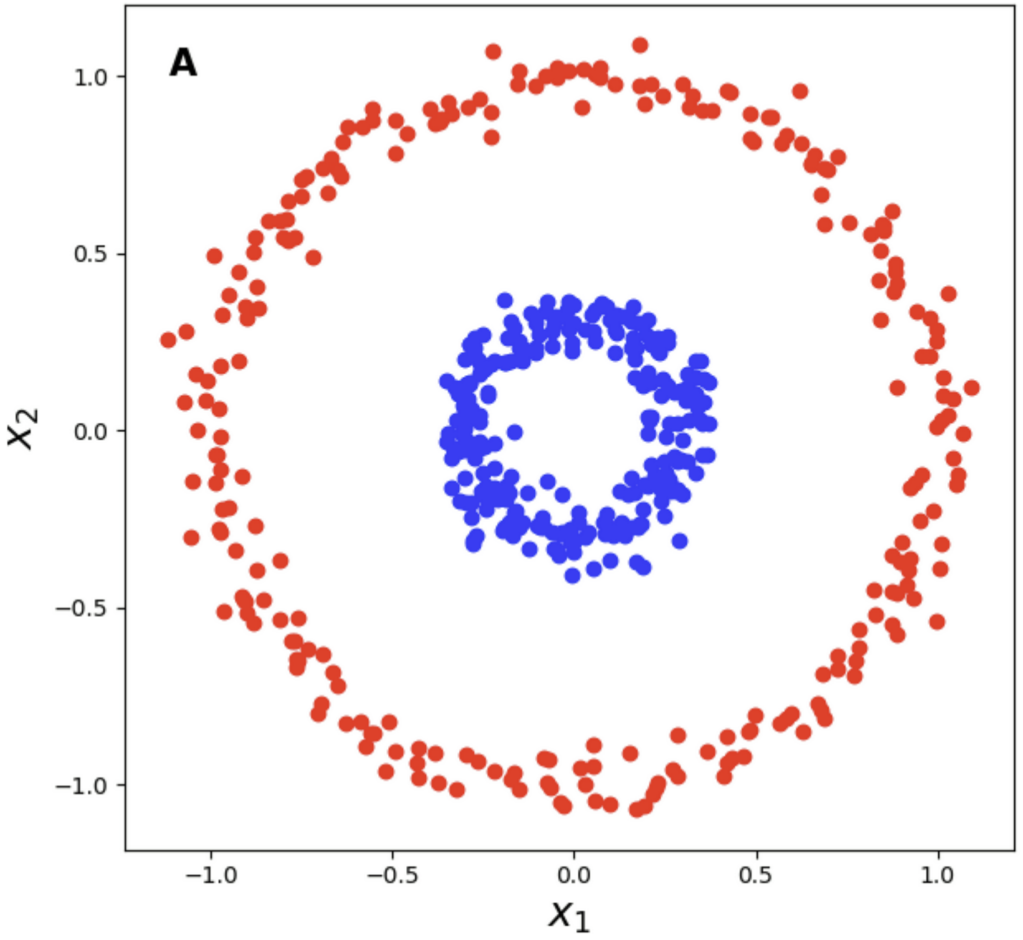
Linear method can not provide good solution

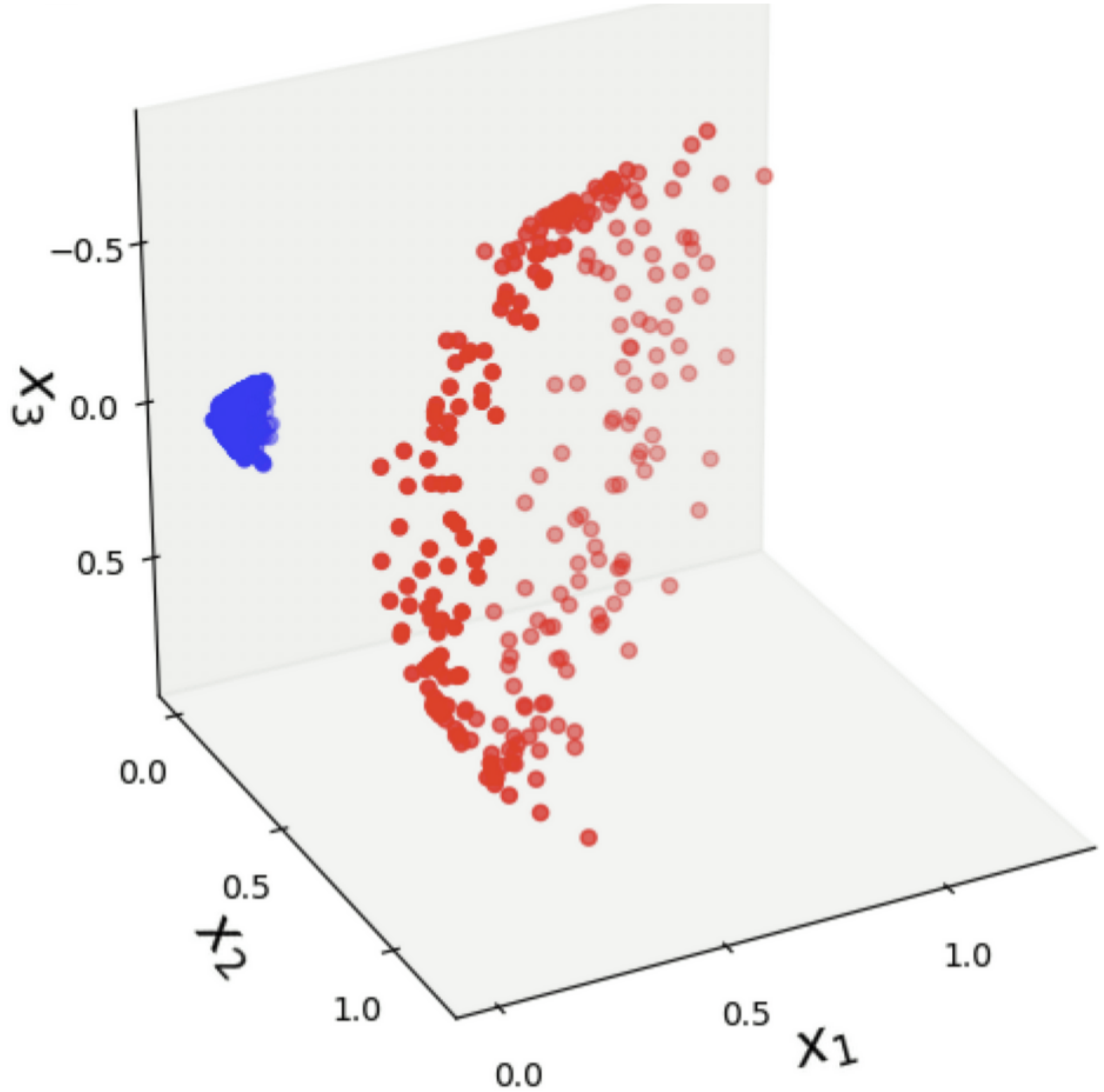


How to deal with non linear boundaries ? *(Classification or regression)*

Linear method can not provide good solution

- If possible add one dimension = add predictor

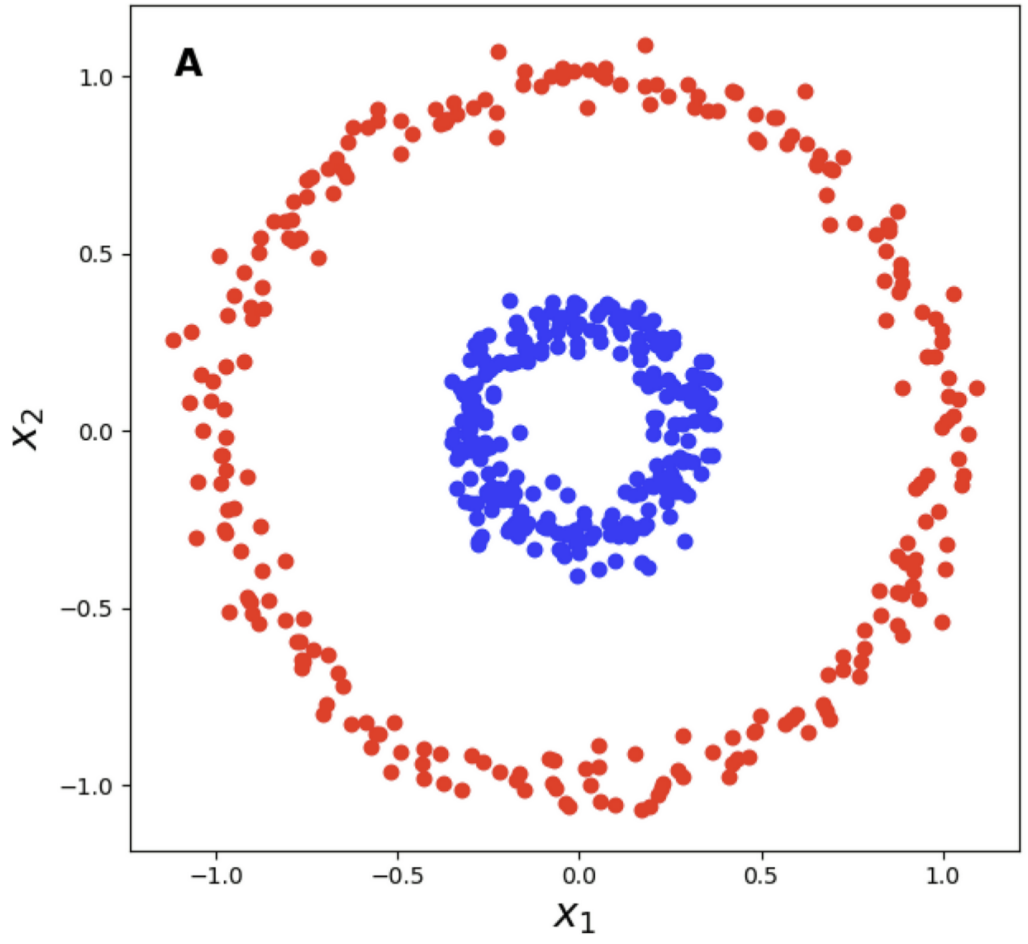


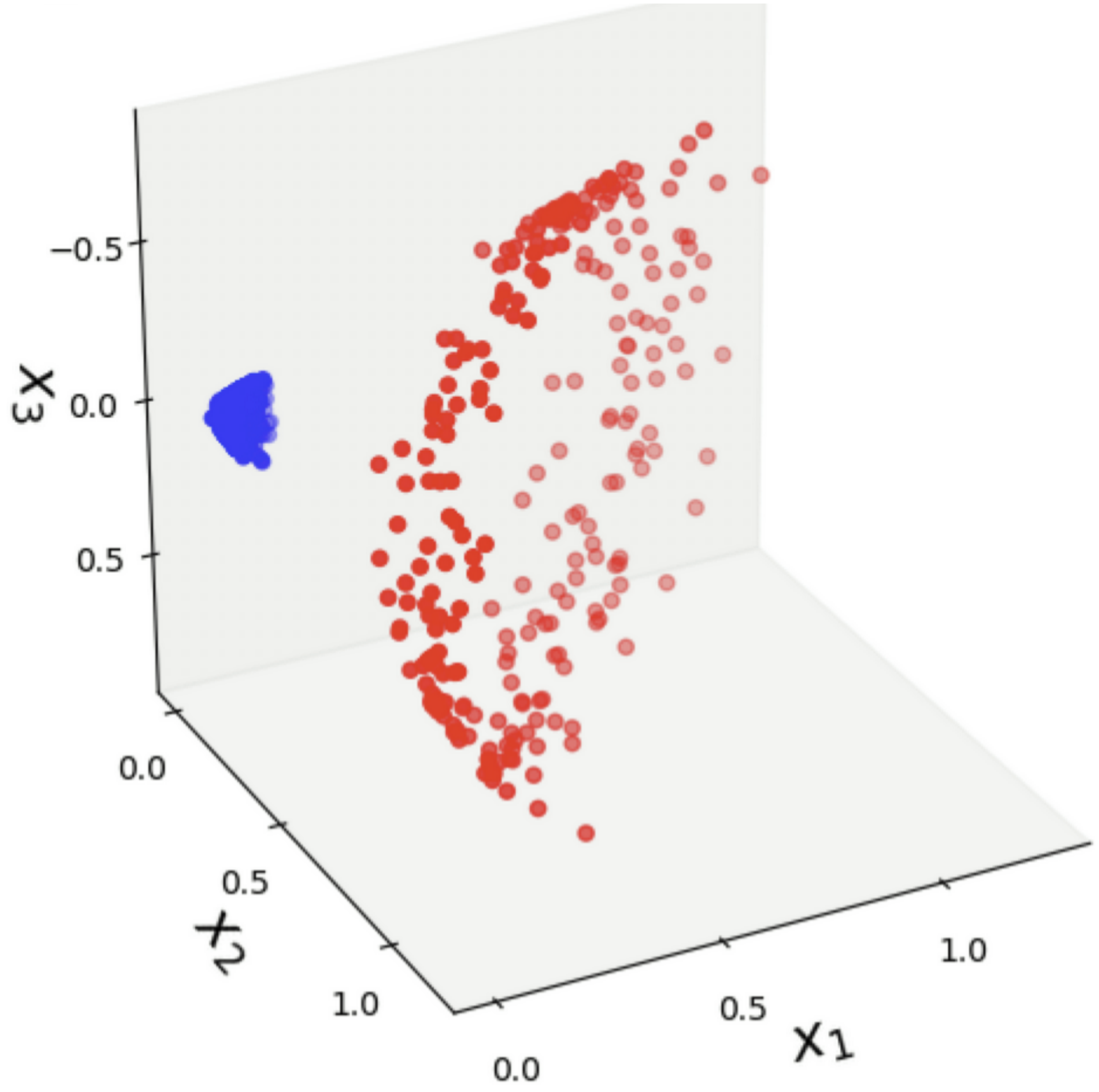


How to deal with non linear boundaries ? *(Classification or regression)*

Linear method can not provide good solution

- If possible add one dimension = add predictor
- Design some features that will present the data differently to let appear a separation plane

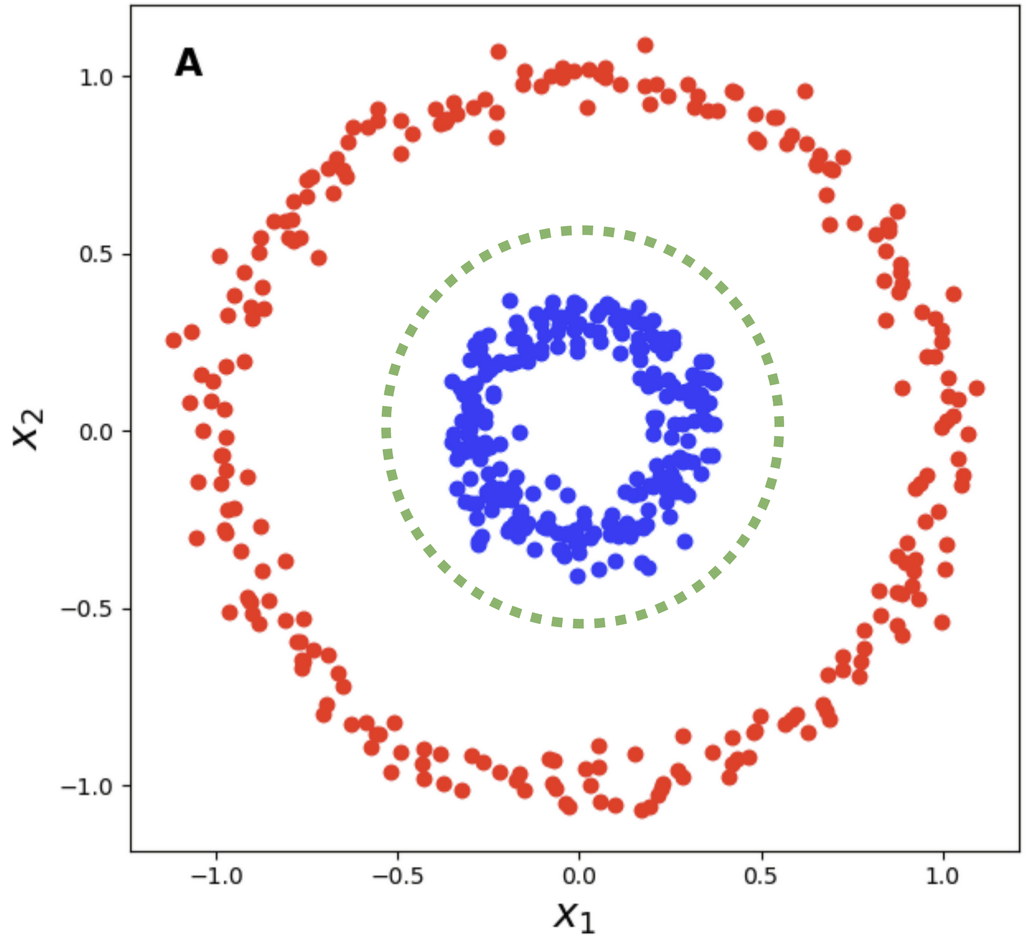




How to deal with non linear boundaries ? *(Classification or regression)*

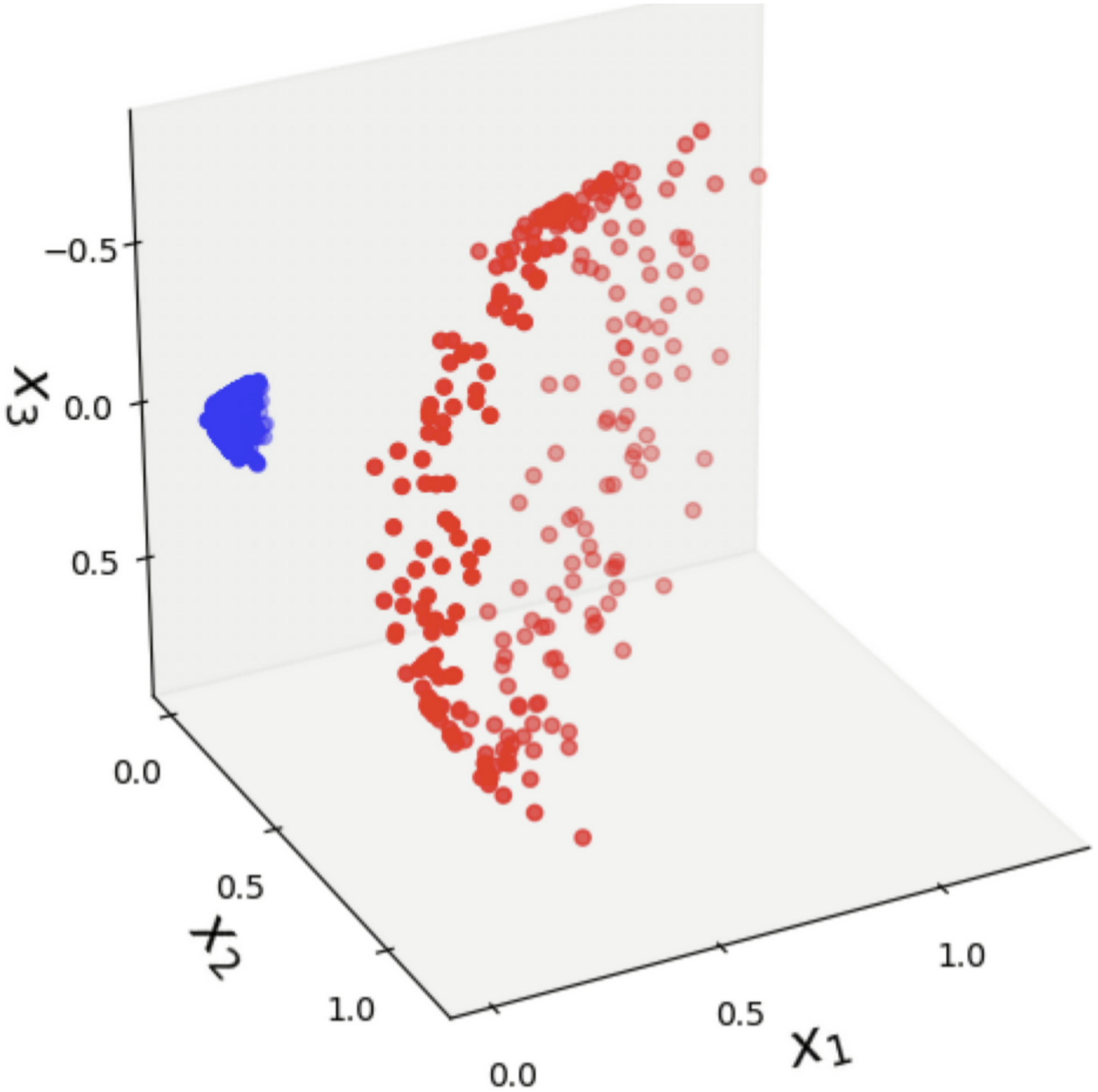
Linear method can not provide good solution

- If possible add one dimension = add predictor
- Design some features that will present the data differently to let appear a separation plane



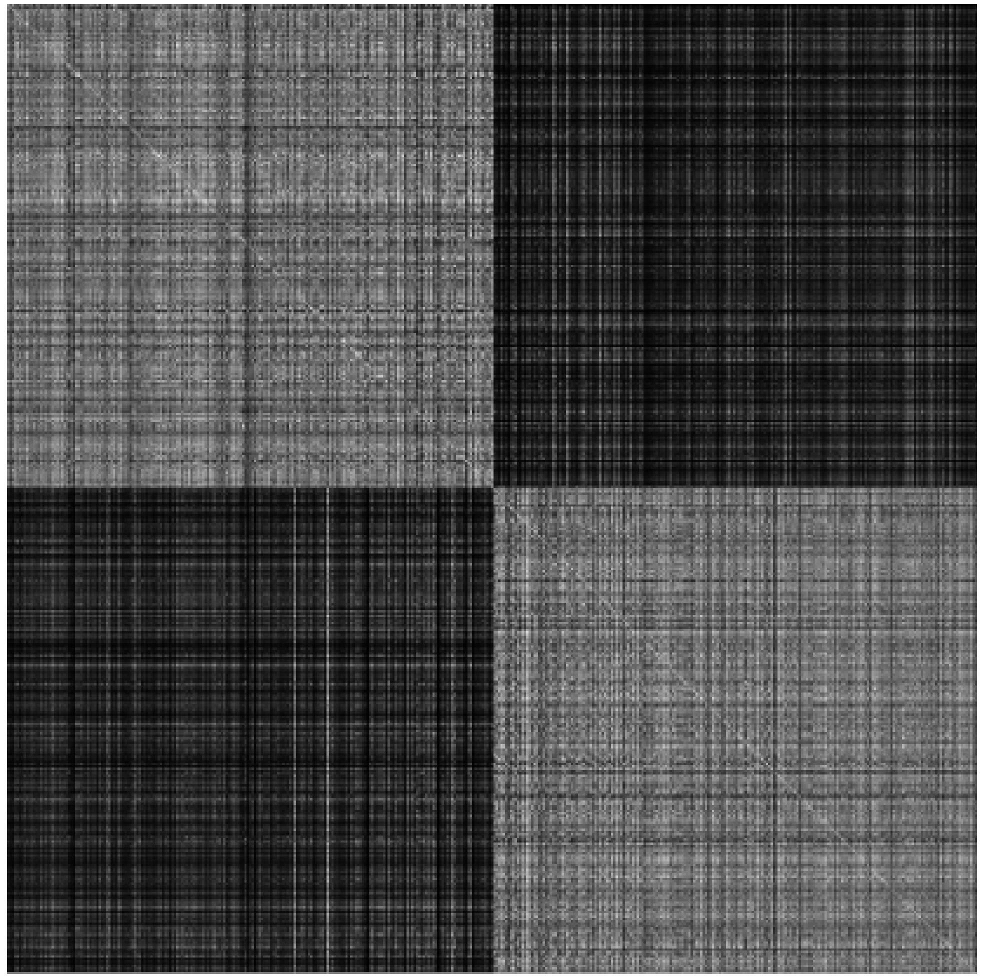
- Use flexible method

  (NN, KNN, k-means, kernel-based methods..)



Recall: Spectral clustering

$\mathbf{K} =$



$= (K(x_i, x_j))_{i,j\in[n]} \in \mathbb{R}^{n\times n}$

Two classes:

- $x_1, \ldots, x_{n/2} \sim \mathcal{N}(\mu, I_p)$
- $x_{n/2+1}, \ldots, x_n \sim \mathcal{N}(-\mu, I_p)$

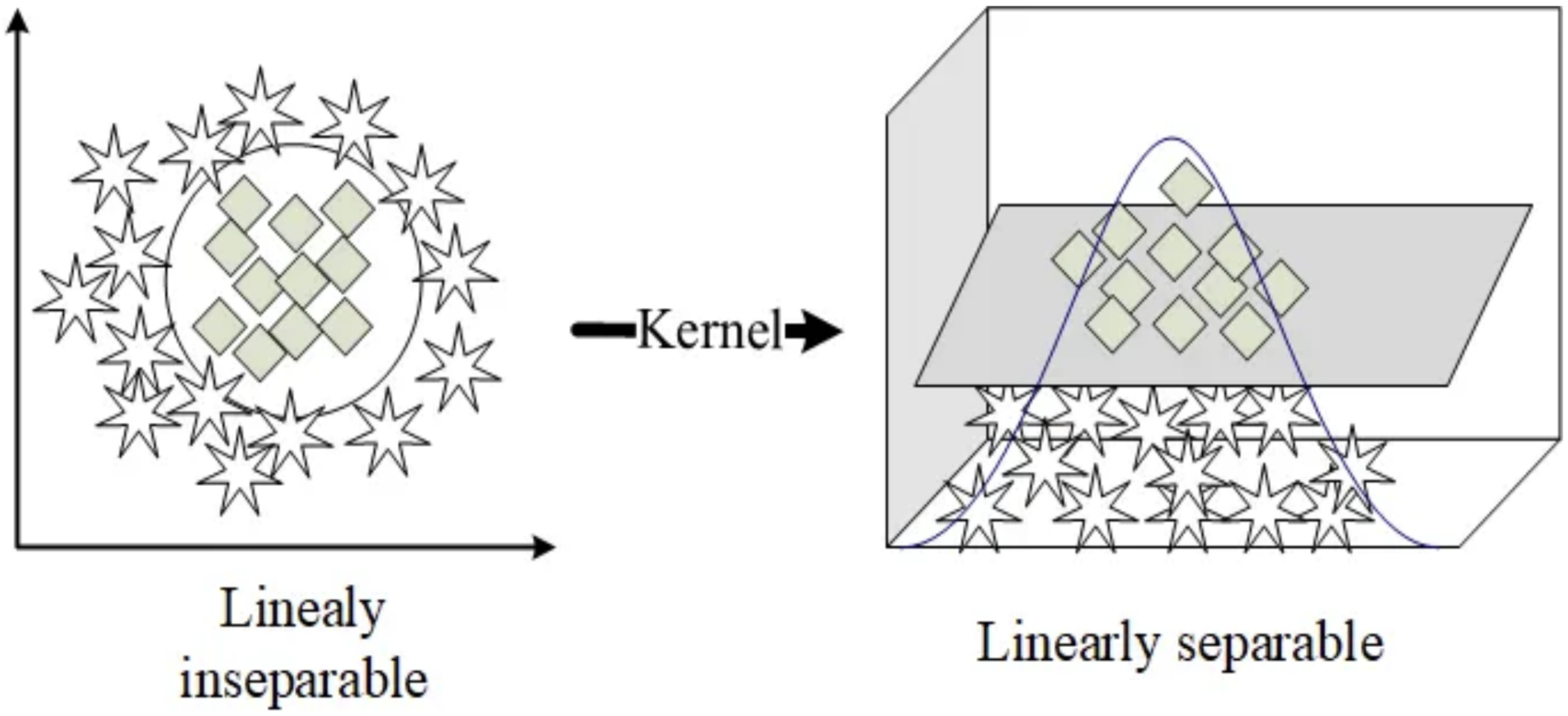
with $\mu = (2, 0, \ldots, 0) \in \mathbb{R}^p$, $n = 500$, $p = 5$.

with for ex. $K(x, y) = e^{-\frac{\|x-y\|^2}{2p}}$ ("Heat kernel")

$\Longrightarrow$ Look for first eigenvectors of K should capture class information

$\mathbf{v}_2 =$

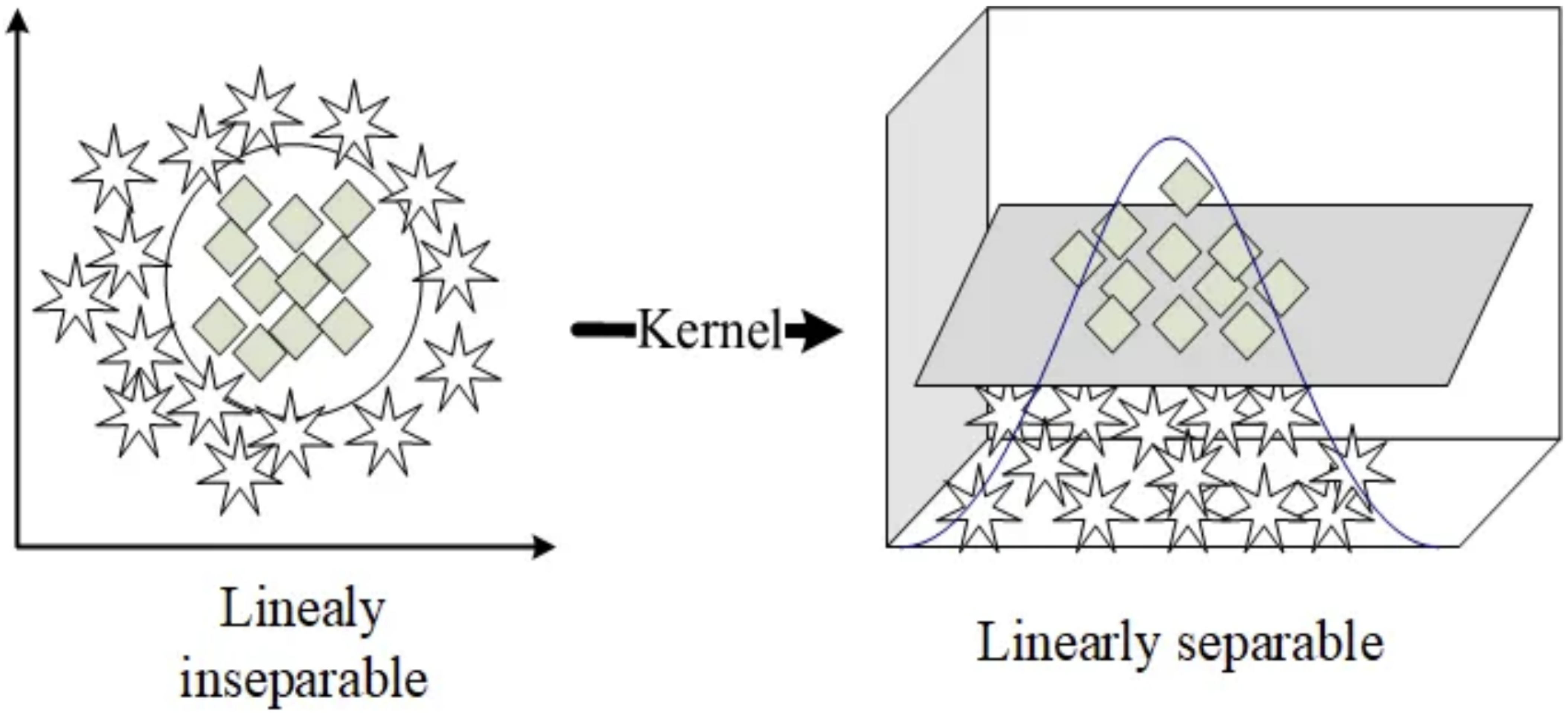
Kernels



Definition: $K : E^2 \to \mathbb{R}$ is a *Mercer Kernel* or a *Positive semi-definite kerneel* iif:

$$\forall x_1, \dots, x_n \in E, \forall c = (c_1, \dots, c_n) \in \mathbb{R}^n : \quad \sum_{i,j=1}^n c_i c_j K(x_i, x_j) \geq 0$$

Kernels



Theorem: Given a Mercer Kernel $K : E^2 \to \mathbb{R}$ there exists:

- a Hilbert space $(H, \langle \cdot | \cdot \rangle)$
- a mapping $\phi : E \to H$

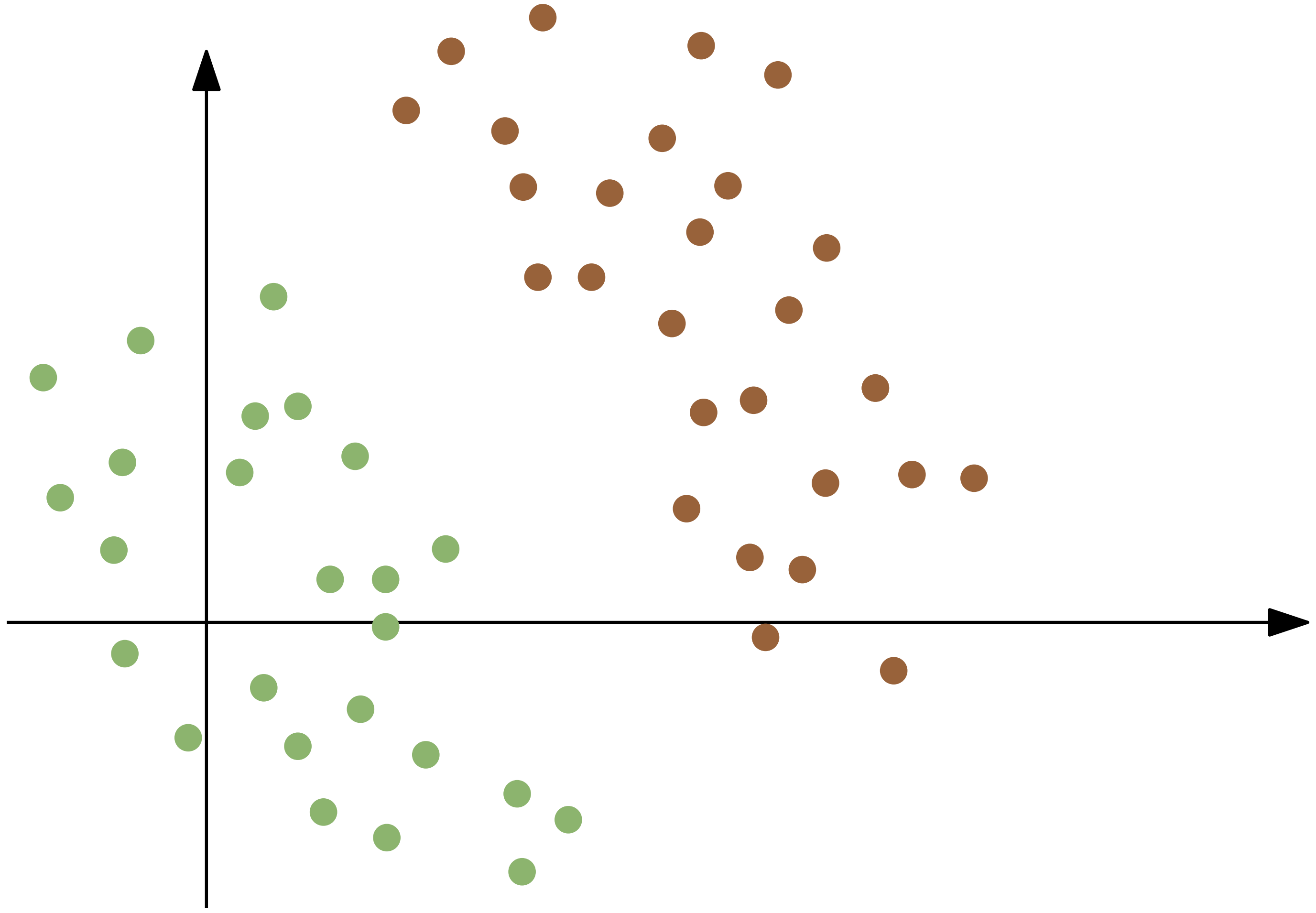
such that:

$$\forall x, y \in E : \quad K(x, y) = \langle \phi(x), \phi(y) \rangle$$

Definition: $K : E^2 \to \mathbb{R}$ is a *Mercer Kernel* or a *Positive semi-definite kerneel* iif:

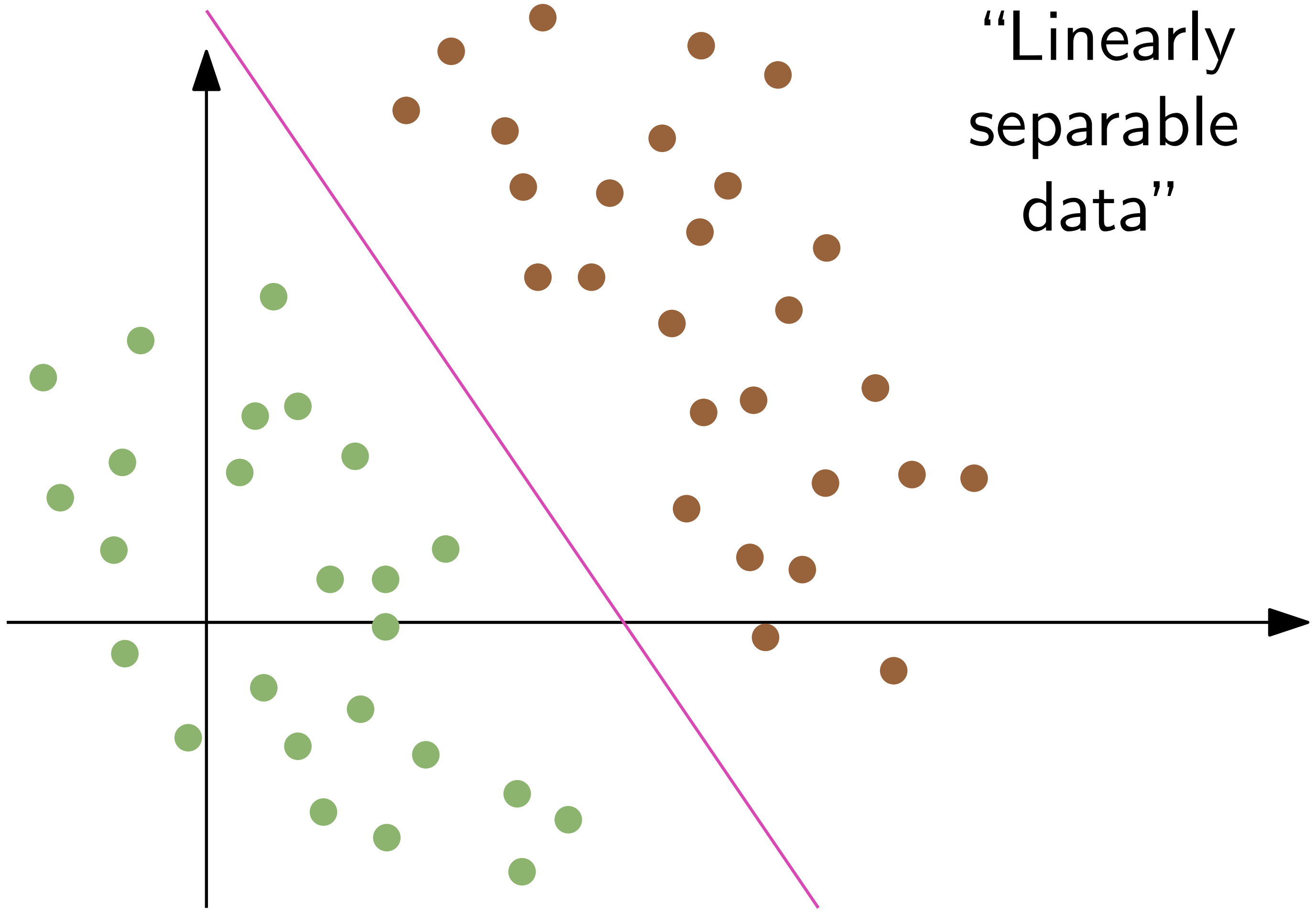
$$\forall x_1, \ldots, x_n \in E, \forall c = (c_1, \ldots, c_n) \in \mathbb{R}^n : \quad \sum_{i,j=1}^n c_i c_j K(x_i, x_j) \geq 0$$

Example of Support vector machines



Example of Support vector machines

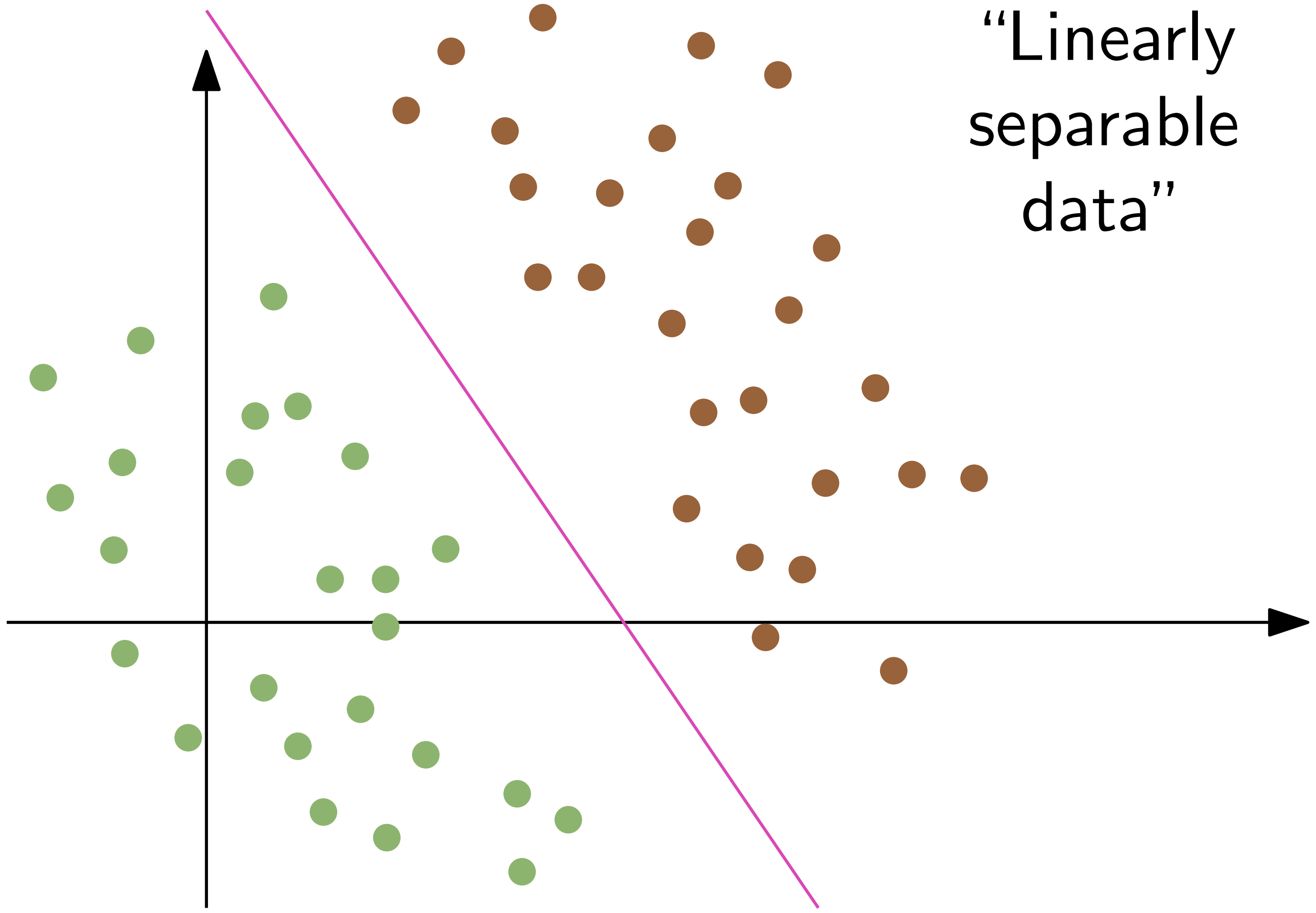
1 - Large (or hard) margin SVM



Example of Support vector machines

1 - Large (or hard) margin SVM

Idea: Look for a linear boundary with largest margin

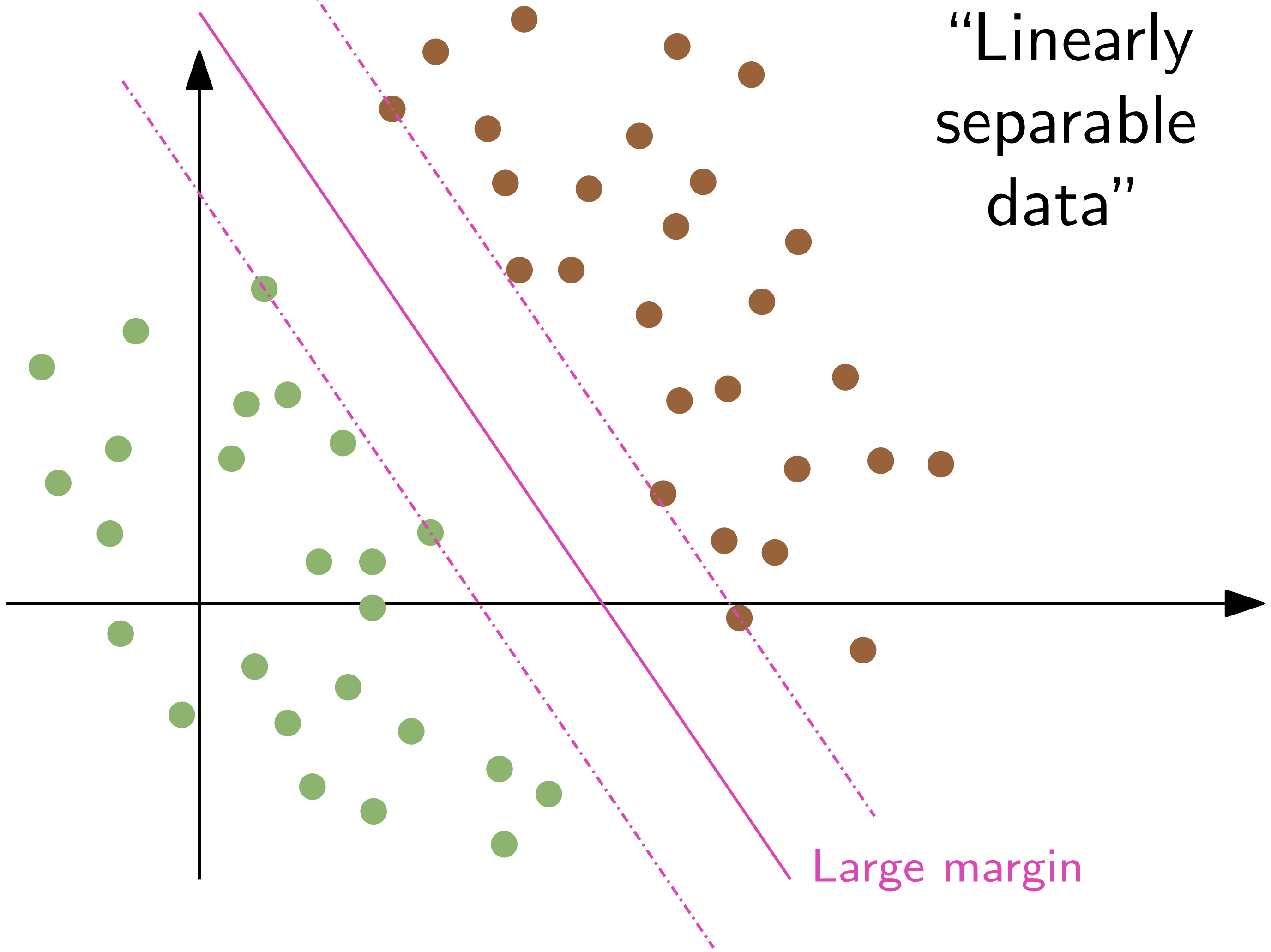


"Linearly separable data"

Example of Support vector machines

1 - Large (or hard) margin SVM

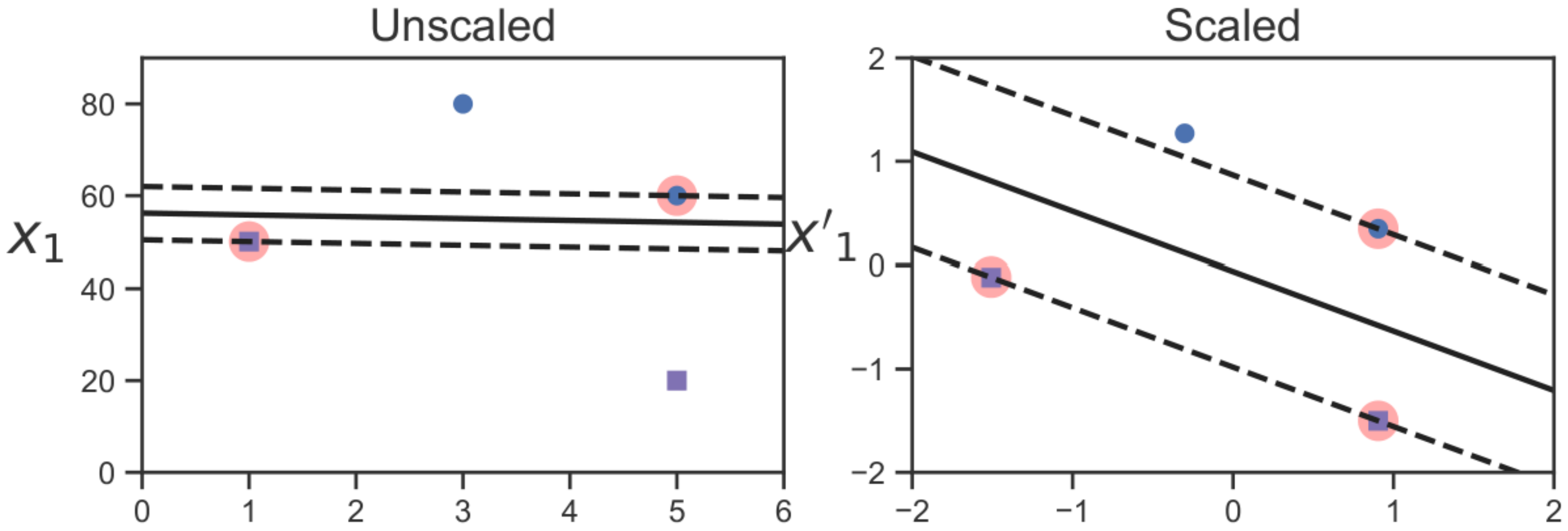
Idea: Look for a linear boundary with largest margin

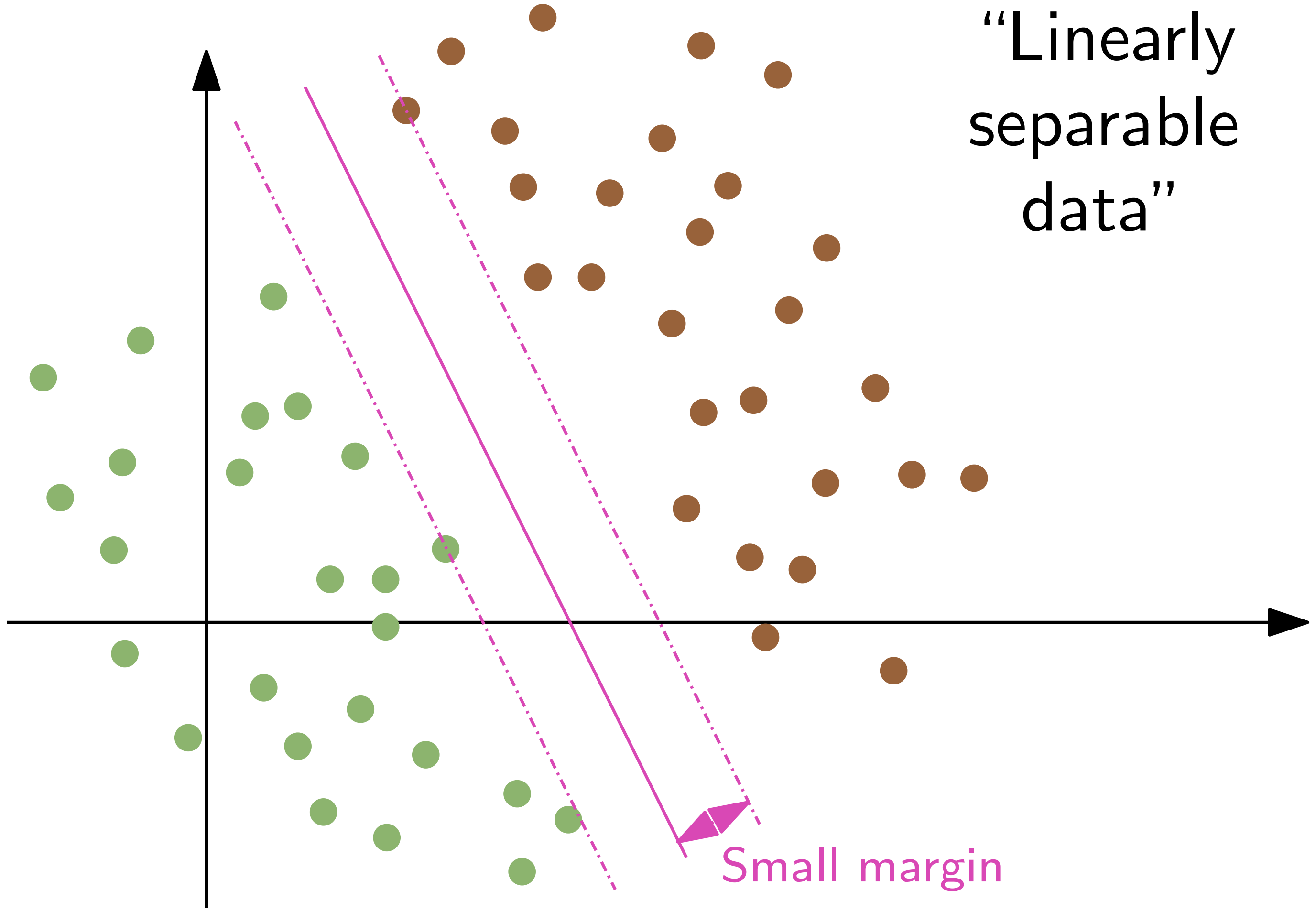


Example of Support vector machines

1 - Large (or hard) margin SVM

Idea: Look for a linear boundary with largest margin

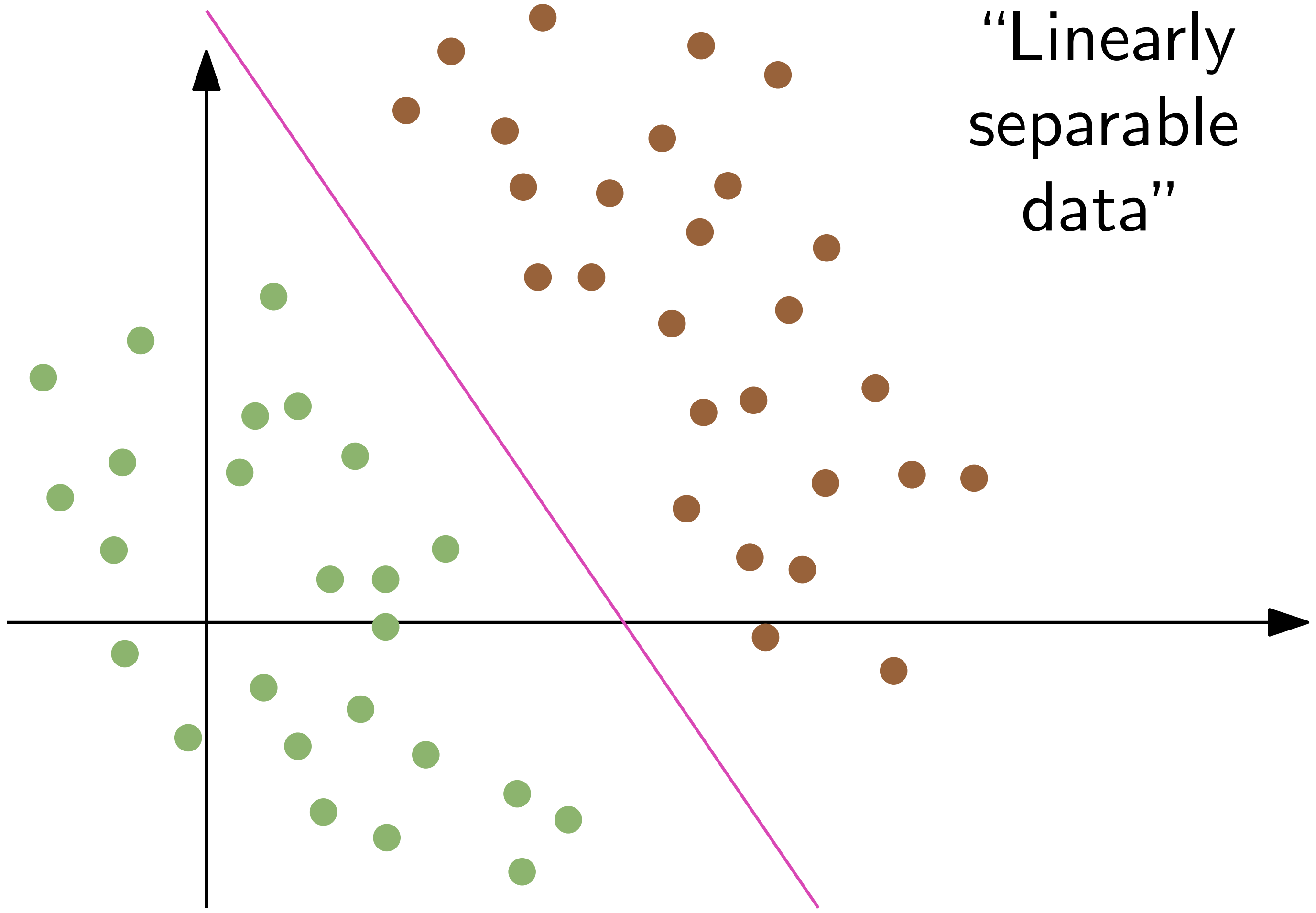




Example of Support vector machines

1 - Large (or hard) margin SVM

Idea: Look for a linear boundary with largest margin

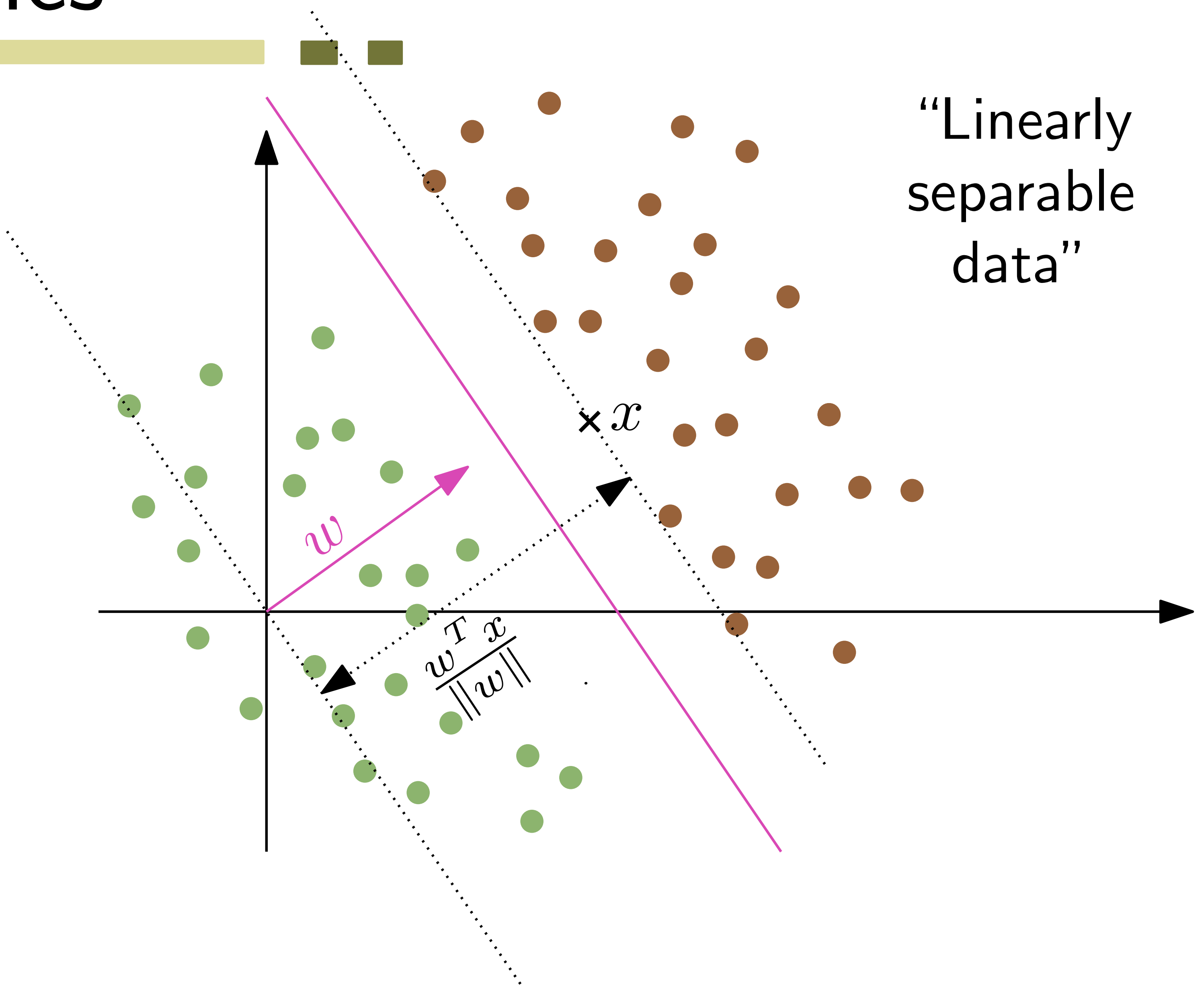


"Linearly separable data"

Example of Support vector machines

1 - Large (or hard) margin SVM

Idea: Look for a linear boundary with largest margin



"Linearly separable data"

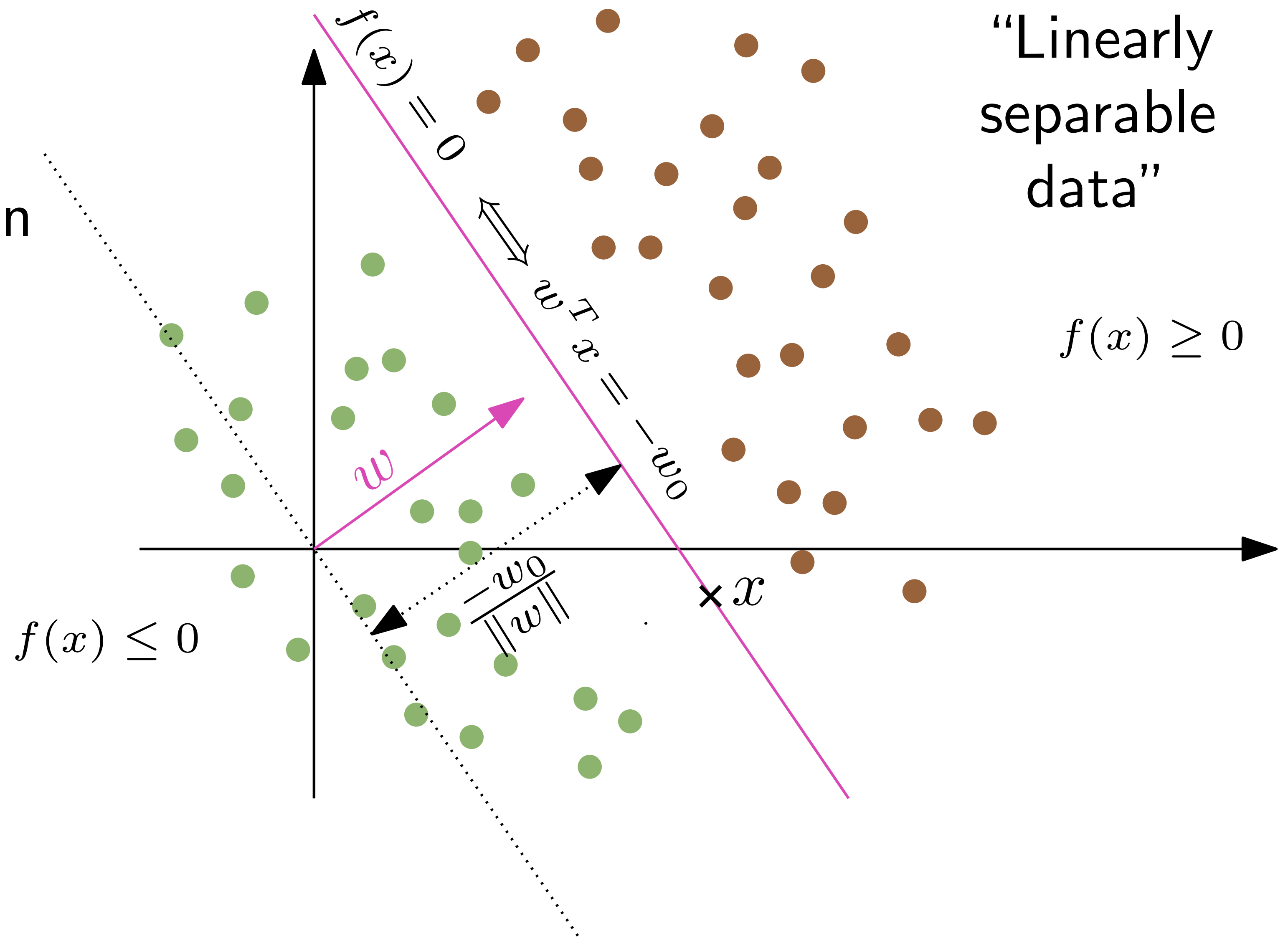
Example of Support vector machines

1 - Large (or hard) margin SVM

Idea: Look for a linear boundary with largest margin

Boundary defined by $w \in \mathbb{R}^p$ and $w_0 \in \mathbb{R}$:

$$f(x) = w^T x + w_0$$



“Linearly separable data”

Example of Support vector machines

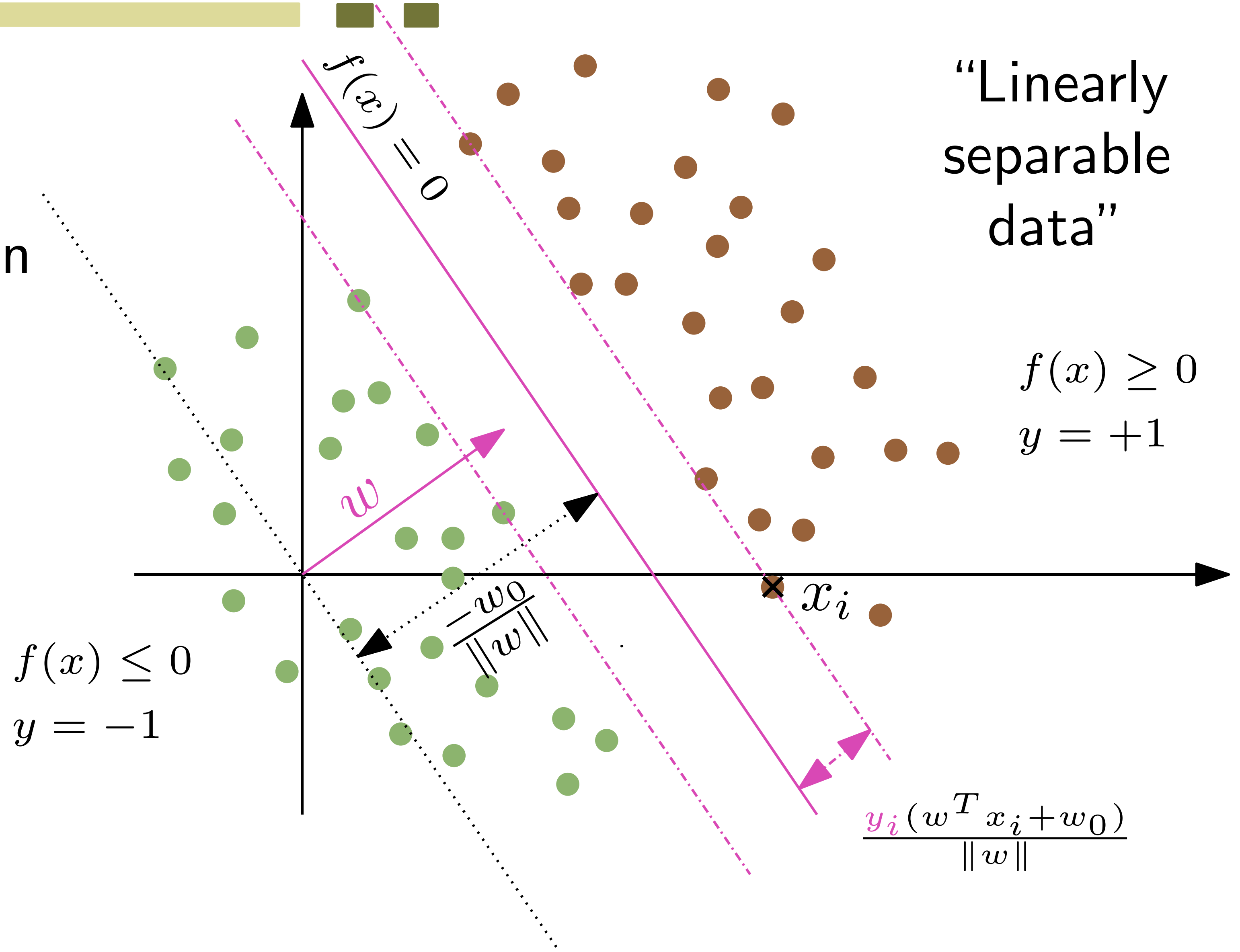
1 - Large (or hard) margin SVM

Idea: Look for a linear boundary with largest margin

Boundary defined by $w \in \mathbb{R}^p$ and $w_0 \in \mathbb{R}$:

$$f(x) = w^T x + w_0$$

$$\text{Margin}(w, w_0) = \min_{i\in[n]} \frac{y_i f(x_i)}{\|w\|}$$



Example of Support vector machines

1 - Large (or hard) margin SVM

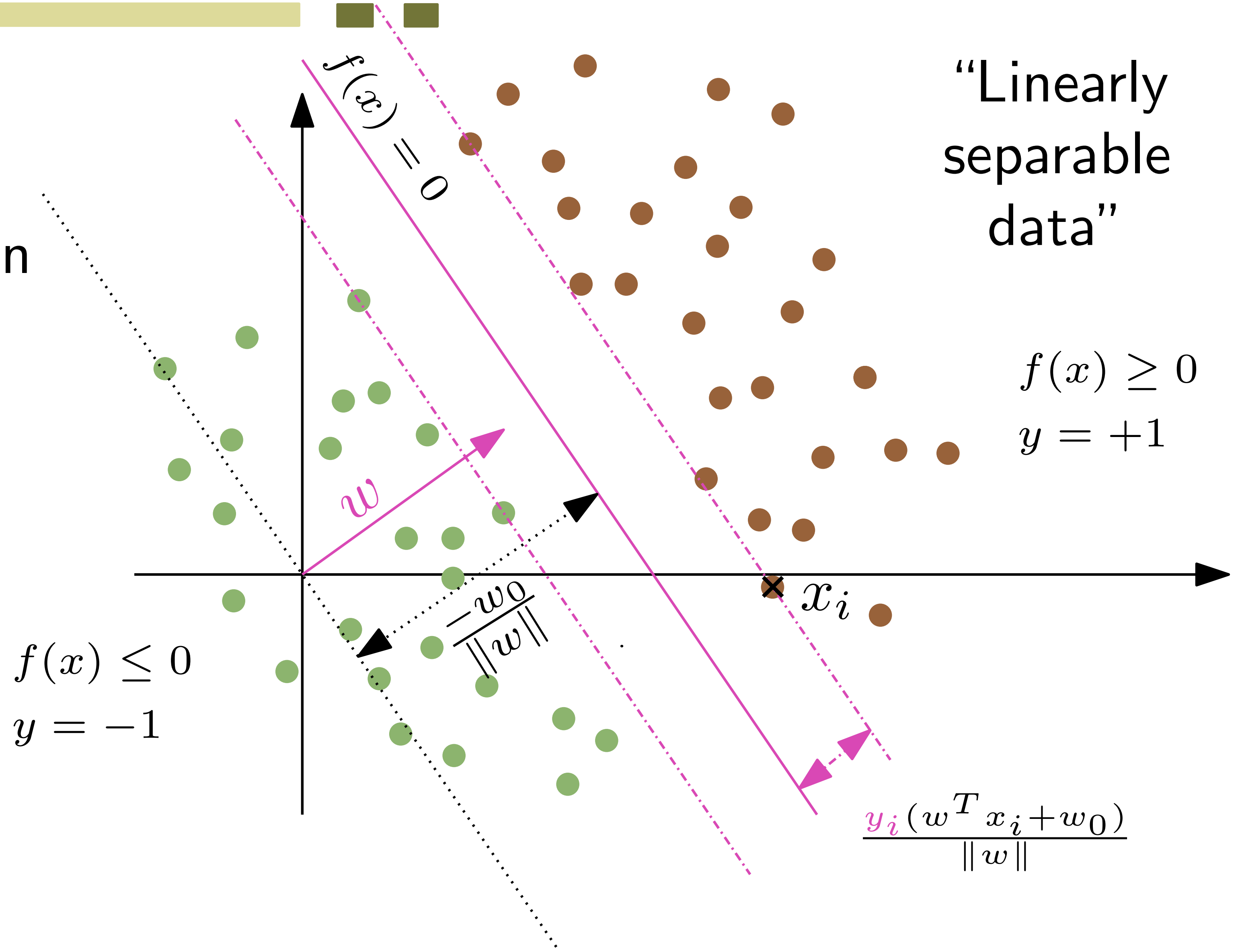
Idea: Look for a linear boundary with largest margin

Boundary defined by $w \in \mathbb{R}^p$ and $w_0 \in \mathbb{R}$:

$$f(x) = w^T x + w_0$$

$$\text{Margin}(w, w_0) = \min_{i \in [n]} \frac{y_i f(x_i)}{\|w\|}$$

$$\underset{w \in \mathbb{R}^p, w_0 \in \mathbb{R}}{\text{Maximize}} \quad \min_{i \in [n]} \frac{y_i (w^T x_i + w_0)}{\|w\|} \quad ?$$



Example of Support vector machines

1 - Large (or hard) margin SVM

Idea: Look for a linear boundary with largest margin

Boundary defined by $w \in \mathbb{R}^p$ and $w_0 \in \mathbb{R}$:

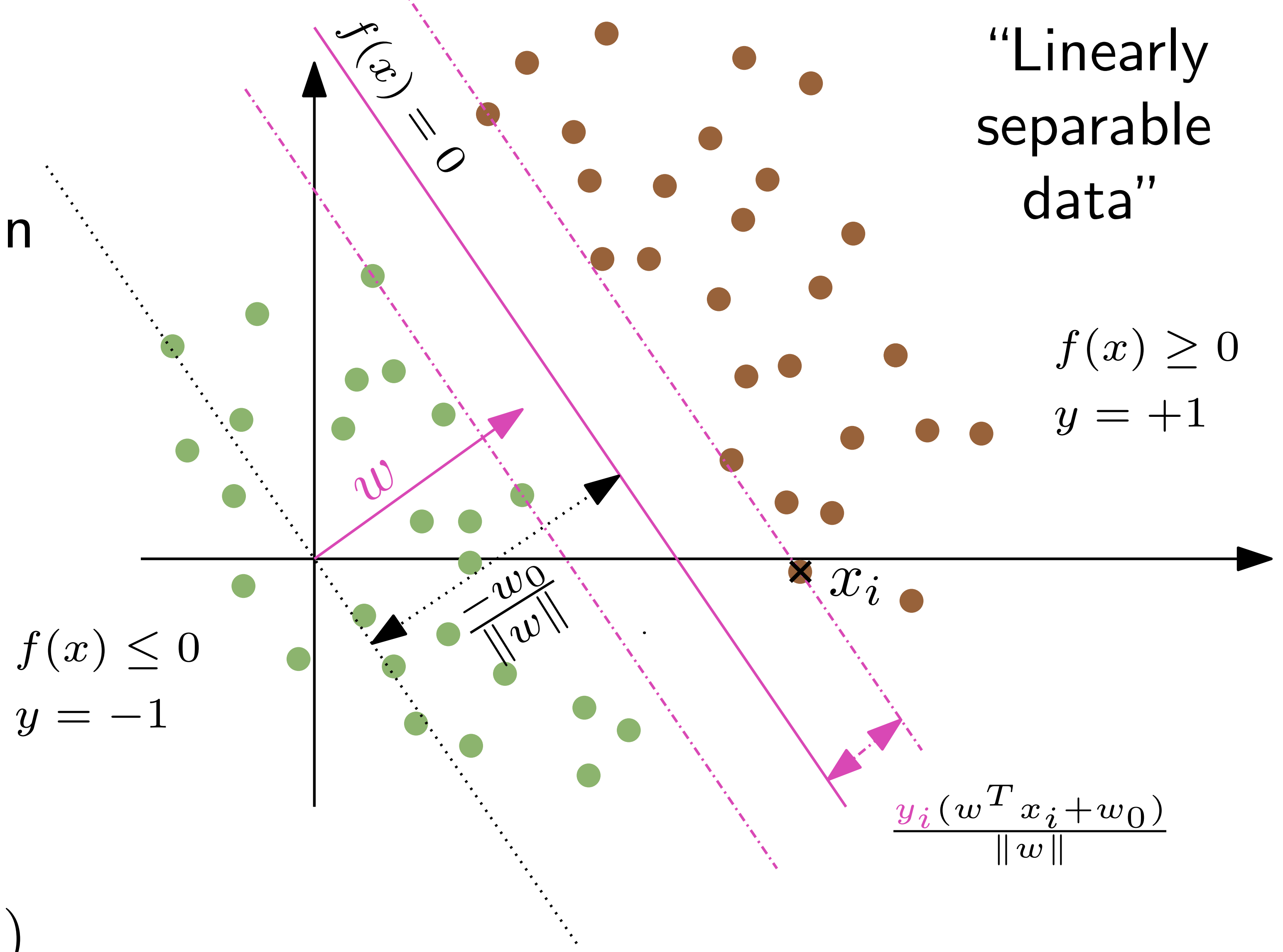
$$f(x) = w^T x + w_0$$

$$\text{Margin}(w, w_0) = \min_{i \in [n]} \frac{y_i f(x_i)}{\|w\|}$$

$$\underset{w \in \mathbb{R}^p, w_0 \in \mathbb{R}}{\text{Maximize}} \quad \min_{i \in [n]} \frac{y_i(w^T x_i + w_0)}{\|w\|} \quad ?$$

Problem: not uniqueness of w, w_0 !

$$\forall \lambda > 0: \ \text{Margin}(\lambda w, \lambda w_0) = \text{Margin}(w, w_0)$$



Example of Support vector machines

1 - Large (or hard) margin SVM

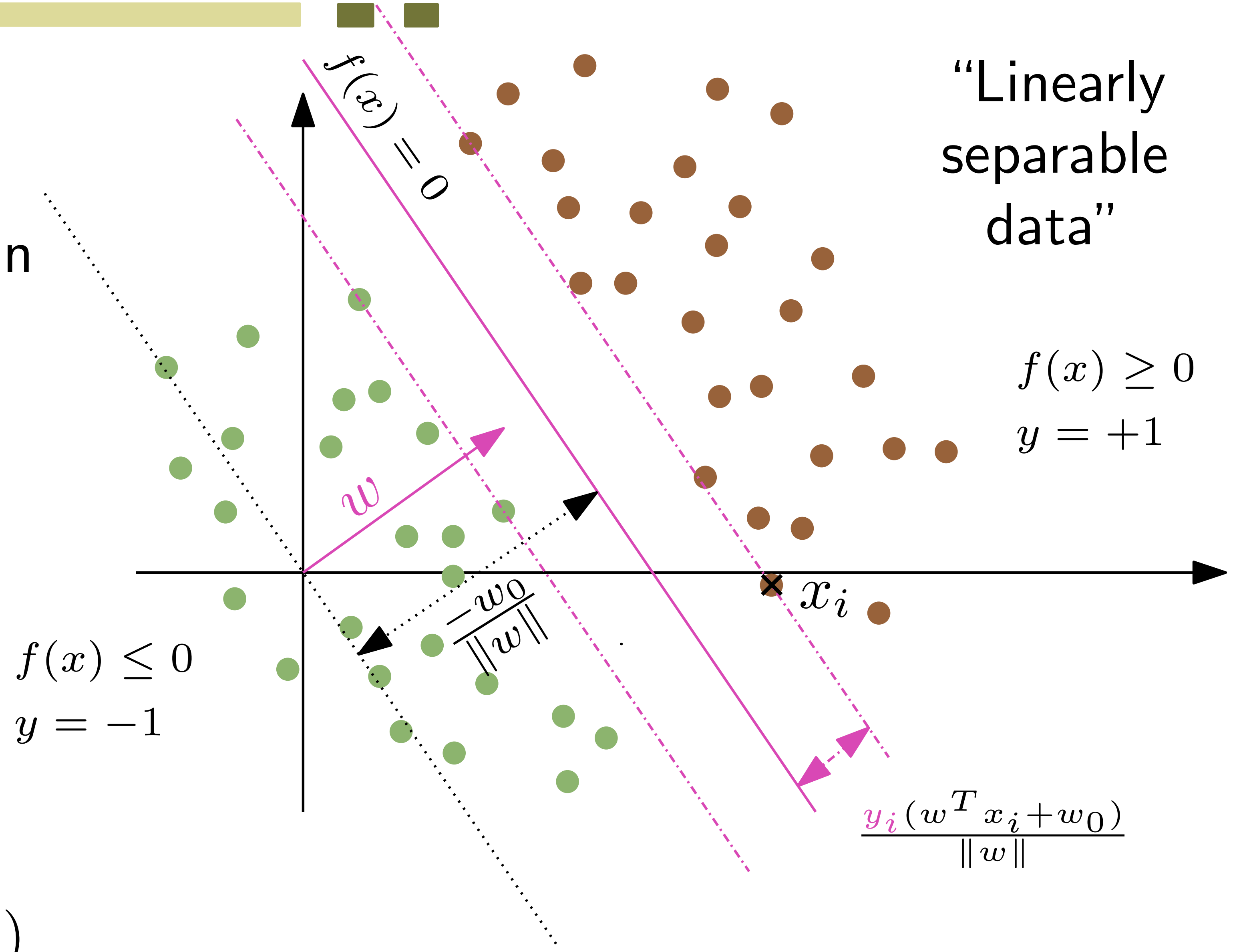
Idea: Look for a linear boundary with largest margin

Boundary defined by $w \in \mathbb{R}^p$ and $w_0 \in \mathbb{R}$:

$$f(x) = w^T x + w_0$$

$$\text{Margin}(w, w_0) = \min_{i \in [n]} \frac{y_i f(x_i)}{\|w\|}$$

$$\underset{w \in \mathbb{R}^p, w_0 \in \mathbb{R}}{\text{Maximize}} \quad \min_{i \in [n]} \frac{y_i(w^T x_i + w_0)}{\|w\|} \quad ?$$



Problem: not uniqueness of w, w_0 !

$$\forall \lambda > 0: \ \text{Margin}(\lambda w, \lambda w_0) = \text{Margin}(w, w_0)$$

Solution: Impose $\forall i \in [n] : \ y_i(w^T x_i + w_0) \geq 1$ then biggest margin: $= \frac{1}{\|w\|}$

Example of Support vector machines

1 - Large (or hard) margin SVM

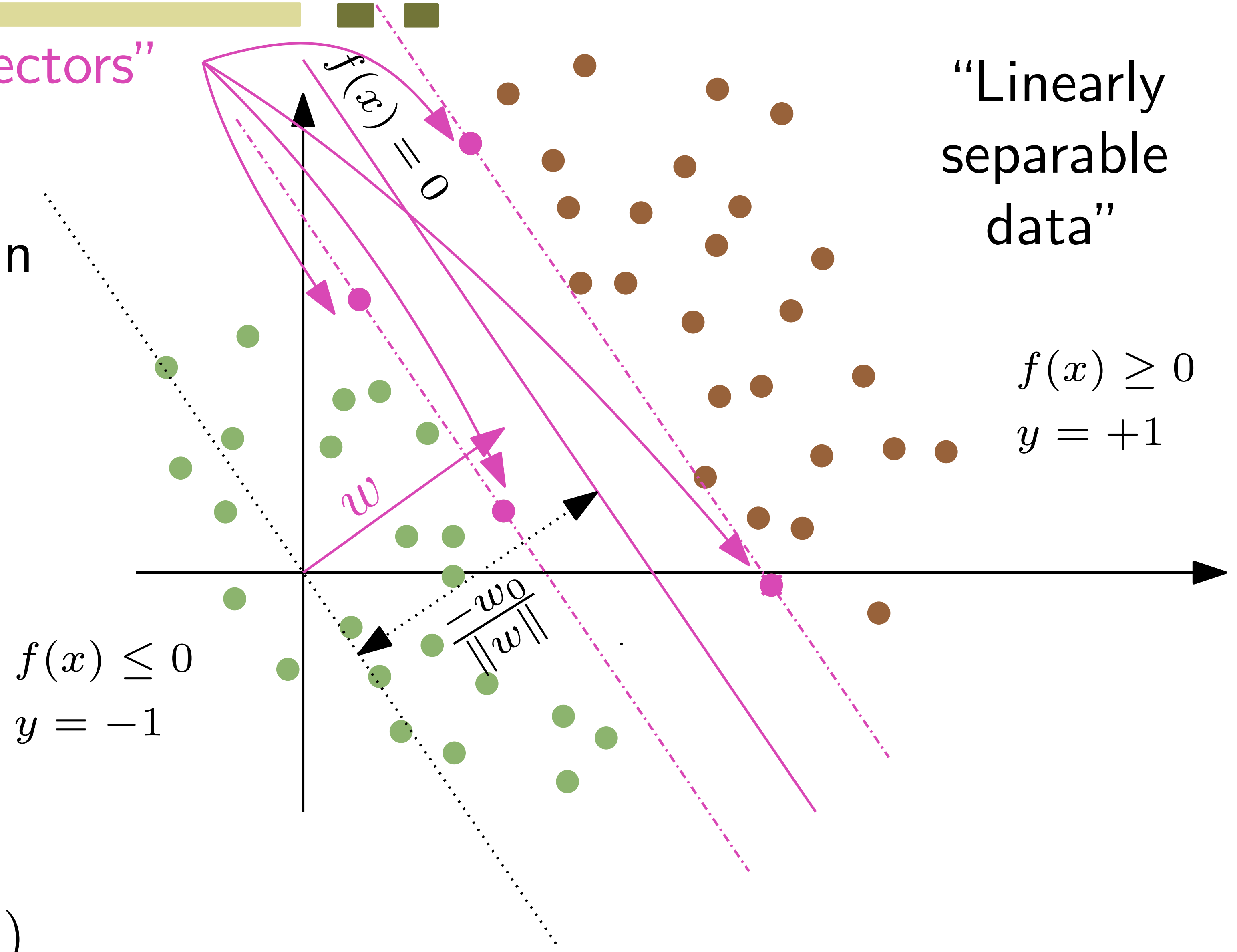
Idea: Look for a linear boundary with largest margin

Boundary defined by $w \in \mathbb{R}^p$ and $w_0 \in \mathbb{R}$:

$$f(x) = w^T x + w_0$$

$$\text{Margin}(w, w_0) = \min_{i\in[n]} \frac{y_i f(x_i)}{\|w\|}$$

$$\underset{w\in\mathbb{R}^p, w_0\in\mathbb{R}}{\text{Maximize}} \quad \min_{i\in[n]} \frac{y_i(w^T x_i + w_0)}{\|w\|} \quad ?$$



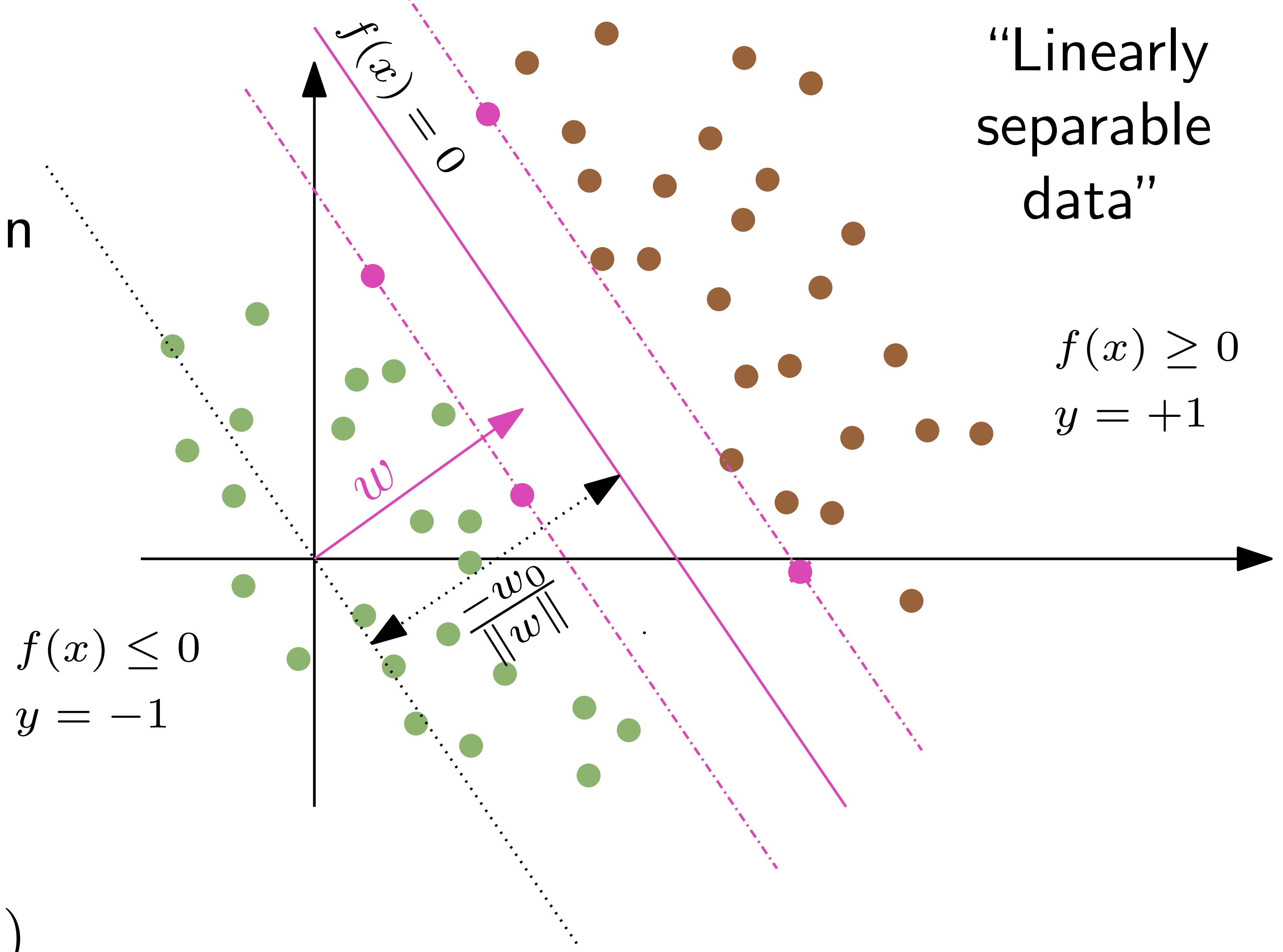
Problem: not uniqueness of w, w_0 !

$$\forall \lambda > 0: \ \text{Margin}(\lambda w, \lambda w_0) = \text{Margin}(w, w_0)$$

Solution: Impose $\forall i \in [n]: \ y_i(w^T x_i + w_0) \geq 1$ then biggest margin: $= \frac{1}{\|w\|}$

Example of Support vector machines

1 - Large (or hard) margin SVM



Idea: Look for a linear boundary with largest margin

Boundary defined by $w \in \mathbb{R}^p$ and $w_0 \in \mathbb{R}$:

$$f(x) = w^T x + w_0$$

$$\text{Margin}(w, w_0) = \min_{i \in [n]} \frac{y_i f(x_i)}{\|w\|}$$

$$\underset{w \in \mathbb{R}^p, w_0 \in \mathbb{R}}{\text{Maximize}} \quad \min_{i \in [n]} \frac{y_i(w^T x_i + w_0)}{\|w\|} \quad ?$$

Problem: not uniqueness of w, w_0 !

$$\forall \lambda > 0: \ \text{Margin}(\lambda w, \lambda w_0) = \text{Margin}(w, w_0)$$

Solution: Impose $\forall i \in [n]: \ y_i(w^T x_i + w_0) \geq 1$ then biggest margin: $= \frac{1}{\|w\|}$

Primal Problem: Maximize $\frac{1}{2}\|w\|^2$ subject to $\forall i \in [n]: \ y_i(w^T x_i + w_0) \geq 1, \quad w \in \mathbb{R}^p, \ w_0 \in \mathbb{R}$.

Large Margin SVM

Primal Problem:

$$\text{Minimize } \frac{1}{2}\|w\|^2 \text{ subject to } \forall i \in [n] : \ y_i(w^T x_i + w_0) \geq 1, \quad w \in \mathbb{R}^p, \ w_0 \in \mathbb{R} \qquad (P)$$

Large Margin SVM

Primal Problem:

$$\text{Minimize } \tfrac{1}{2}\|w\|^2 \text{ subject to } \forall i \in [n]: \ y_i(w^T x_i + w_0) \geq 1, \quad w \in \mathbb{R}^p,\ w_0 \in \mathbb{R} \qquad (P)$$

Lagrangian: $L(w, w_0, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^n \alpha_i \left(y_i(w^T x_i + w_0) - 1\right), \quad w \in \mathbb{R}^p, w_0 \in \mathbb{R}, \alpha \in \mathbb{R}_+^p$

Large Margin SVM

Primal Problem:

$$\text{Minimize } \tfrac{1}{2}\|w\|^2 \text{ subject to } \forall i \in [n]: \; y_i(w^T x_i + w_0) \geq 1, \quad w \in \mathbb{R}^p,\ w_0 \in \mathbb{R} \qquad (P)$$

Lagrangian: $L(w, w_0, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^n \alpha_i \left(y_i(w^T x_i + w_0) - 1\right), \quad w \in \mathbb{R}^p, w_0 \in \mathbb{R}, \alpha \in \mathbb{R}^p_+$

$$\hat{w}, \hat{w}_0 \text{ solve } (P) \quad \Longleftrightarrow \quad \frac{1}{2}\|\hat{w}\|^2 = \max_{\alpha \in \mathbb{R}^n_+} L(\hat{w}, \hat{w}_0, \alpha)$$

Large Margin SVM

Primal Problem:

$$\text{Minimize } \tfrac{1}{2}\|w\|^2 \text{ subject to } \forall i \in [n]: \; y_i(w^T x_i + w_0) \geq 1, \quad w \in \mathbb{R}^p,\, w_0 \in \mathbb{R} \qquad (P)$$

Lagrangian: $L(w, w_0, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^n \alpha_i \left(y_i(w^T x_i + w_0) - 1 \right), \quad w \in \mathbb{R}^p, w_0 \in \mathbb{R}, \alpha \in \mathbb{R}^p_+$

$$\hat{w}, \hat{w}_0 \text{ solve } (P) \quad \Longleftrightarrow \quad \frac{1}{2}\|\hat{w}\|^2 = \max_{\alpha \in \mathbb{R}^n_+} L(\hat{w}, \hat{w}_0, \alpha)$$

(Because if $\exists i \in [n]$ s.t. $y_i(w^T x_i + w_0) < 1$ then $\max_{\alpha_i \geq 0} -\alpha_i(y_i(w^T x_i + w_0) - 1)) = +\infty$)

Large Margin SVM

Primal Problem:

$$\text{Minimize } \frac{1}{2}\|w\|^2 \text{ subject to } \forall i \in [n] : \; y_i(w^T x_i + w_0) \geq 1, \quad w \in \mathbb{R}^p,\; w_0 \in \mathbb{R} \qquad (P)$$

Lagrangian: $L(w, w_0, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^n \alpha_i \left(y_i(w^T x_i + w_0) - 1\right), \quad w \in \mathbb{R}^p, w_0 \in \mathbb{R}, \alpha \in \mathbb{R}_+^p$

$$\hat{w}, \hat{w}_0 \text{ solve } (P) \quad \Longleftrightarrow \quad \frac{1}{2}\|\hat{w}\|^2 = \max_{\alpha \in \mathbb{R}_+^n} L(\hat{w}, \hat{w}_0, \alpha)$$

Thus: $\hat{w}, \hat{w}_0$ solve $(P) \quad \Longleftrightarrow \quad \hat{w}, \hat{w}_0 = \underset{w \in \mathbb{R}^p, w_0 \in \mathbb{R}}{\text{Argmin}} \; \max_{\alpha \in \mathbb{R}_+^n} L(w, w_0, \alpha)$

Large Margin SVM

Primal Problem:

$$\text{Minimize } \tfrac{1}{2}\|w\|^2 \text{ subject to } \forall i \in [n] : \ y_i(w^T x_i + w_0) \geq 1, \quad w \in \mathbb{R}^p,\ w_0 \in \mathbb{R} \qquad (P)$$

Lagrangian: $L(w, w_0, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^n \alpha_i \left(y_i(w^T x_i + w_0) - 1 \right), \quad w \in \mathbb{R}^p, w_0 \in \mathbb{R}, \alpha \in \mathbb{R}_+^p$

$$\hat{w}, \hat{w}_0 \text{ solve } (P) \iff \tfrac{1}{2}\|\hat{w}\|^2 = \max_{\alpha \in \mathbb{R}_+^n} L(\hat{w}, \hat{w}_0, \alpha)$$

Thus: $\hat{w}, \hat{w}_0$ solve $(P) \iff \hat{w}, \hat{w}_0 = \underset{w \in \mathbb{R}^p, w_0 \in \mathbb{R}}{\text{Argmin}} \ \max_{\alpha \in \mathbb{R}_+^n} L(w, w_0, \alpha)$

Optimization theory Theorem: There exist $\hat{w}, \hat{w}_0, \hat{\alpha}$ s.t. (denoting $f : x \mapsto \hat{w}^T x + \hat{w}_0$):

- $\displaystyle \min_{w \in \mathbb{R}^p, w_0 \in \mathbb{R}} \ \max_{\alpha \in \mathbb{R}_+^n} L(w, w_0, \alpha) = L(\hat{w}, \hat{w}_0, \hat{\alpha}) = \max_{\alpha \in \mathbb{R}_+^n} \ \min_{w \in \mathbb{R}^p, w_0 \in \mathbb{R}} L(w, w_0, \alpha)$

Large Margin SVM

Primal Problem:

$$\text{Minimize } \tfrac{1}{2}\|w\|^2 \text{ subject to } \forall i \in [n] : \ y_i(w^T x_i + w_0) \geq 1, \quad w \in \mathbb{R}^p,\ w_0 \in \mathbb{R} \qquad (P)$$

Lagrangian: $L(w, w_0, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^n \alpha_i \left(y_i(w^T x_i + w_0) - 1\right), \quad w \in \mathbb{R}^p, w_0 \in \mathbb{R}, \alpha \in \mathbb{R}_+^p$

$$\hat{w}, \hat{w}_0 \text{ solve } (P) \iff \tfrac{1}{2}\|\hat{w}\|^2 = \max_{\alpha \in \mathbb{R}_+^n} L(\hat{w}, \hat{w}_0, \alpha)$$

Thus: $\hat{w}, \hat{w}_0$ solve $(P) \iff \hat{w}, \hat{w}_0 = \underset{w \in \mathbb{R}^p, w_0 \in \mathbb{R}}{\text{Argmin}} \max_{\alpha \in \mathbb{R}_+^n} L(w, w_0, \alpha)$

Optimization theory Theorem: There exist $\hat{w}, \hat{w}_0, \hat{\alpha}$ s.t. (denoting $f : x \mapsto \hat{w}^T x + \hat{w}_0$):

- $\min_{w \in \mathbb{R}^p, w_0 \in \mathbb{R}} \max_{\alpha \in \mathbb{R}_+^n} L(w, w_0, \alpha) = L(\hat{w}, \hat{w}_0, \hat{\alpha}) = \max_{\alpha \in \mathbb{R}_+^n} \min_{w \in \mathbb{R}^p, w_0 \in \mathbb{R}} L(w, w_0, \alpha)$

Large Margin SVM

Primal Problem:

$$\text{Minimize } \tfrac{1}{2}\|w\|^2 \text{ subject to } \forall i \in [n]:\ y_i(w^T x_i + w_0) \geq 1, \quad w \in \mathbb{R}^p,\ w_0 \in \mathbb{R} \qquad (P)$$

Lagrangian: $L(w, w_0, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^n \alpha_i \left(y_i(w^T x_i + w_0) - 1 \right), \quad w \in \mathbb{R}^p, w_0 \in \mathbb{R}, \alpha \in \mathbb{R}_+^p$

$$\hat{w}, \hat{w}_0 \text{ solve } (P) \iff \frac{1}{2}\|\hat{w}\|^2 = \max_{\alpha \in \mathbb{R}_+^n} L(\hat{w}, \hat{w}_0, \alpha)$$

Thus: $\hat{w}, \hat{w}_0$ solve $(P) \iff \hat{w}, \hat{w}_0 = \underset{w \in \mathbb{R}^p, w_0 \in \mathbb{R}}{\text{Argmin}} \max_{\alpha \in \mathbb{R}_+^n} L(w, w_0, \alpha)$

Optimization theory Theorem: There exist $\hat{w}, \hat{w}_0, \hat{\alpha}$ s.t. (denoting $f : x \mapsto \hat{w}^T x + \hat{w}_0$):

- $\min_{w \in \mathbb{R}^p, w_0 \in \mathbb{R}} \max_{\alpha \in \mathbb{R}_+^n} L(w, w_0, \alpha) = L(\hat{w}, \hat{w}_0, \hat{\alpha}) = \max_{\alpha \in \mathbb{R}_+^n} \min_{w \in \mathbb{R}^p, w_0 \in \mathbb{R}} L(w, w_0, \alpha) \qquad (= \frac{1}{2}\|\hat{w}\|^2)$

Large Margin SVM

Primal Problem:

$$\text{Minimize } \tfrac{1}{2}\|w\|^2 \text{ subject to } \forall i \in [n] : \; y_i(w^T x_i + w_0) \geq 1, \quad w \in \mathbb{R}^p,\, w_0 \in \mathbb{R} \qquad (P)$$

Lagrangian: $L(w, w_0, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^n \alpha_i \left(y_i(w^T x_i + w_0) - 1\right), \quad w \in \mathbb{R}^p, w_0 \in \mathbb{R}, \alpha \in \mathbb{R}^p_+$

$$\hat{w}, \hat{w}_0 \text{ solve } (P) \iff \tfrac{1}{2}\|\hat{w}\|^2 = \max_{\alpha \in \mathbb{R}^n_+} L(\hat{w}, \hat{w}_0, \alpha)$$

Thus: $\hat{w}, \hat{w}_0$ solve $(P) \iff \hat{w}, \hat{w}_0 = \underset{w \in \mathbb{R}^p, w_0 \in \mathbb{R}}{\text{Argmin}} \max_{\alpha \in \mathbb{R}^n_+} L(w, w_0, \alpha)$

Optimization theory Theorem: There exist $\hat{w}, \hat{w}_0, \hat{\alpha}$ s.t. (denoting $f : x \mapsto \hat{w}^T x + \hat{w}_0$):

- $\min_{w \in \mathbb{R}^p, w_0 \in \mathbb{R}} \max_{\alpha \in \mathbb{R}^n_+} L(w, w_0, \alpha) = L(\hat{w}, \hat{w}_0, \hat{\alpha}) = \max_{\alpha \in \mathbb{R}^n_+} \min_{w \in \mathbb{R}^p, w_0 \in \mathbb{R}} L(w, w_0, \alpha) \qquad (= \frac{1}{2}\|\hat{w}\|^2)$
- KKT conditions: $\frac{\partial L}{\partial w}\Big|_{\hat{w}, \hat{w}_0, \hat{\alpha}} = 0, \qquad \frac{\partial L}{\partial w_0}\Big|_{\hat{w}, \hat{w}_0, \hat{\alpha}} = 0,$
  $\forall i \in [n] : \quad \hat{\alpha}_i (f(x_i) - 1) y_i = 0, \quad \hat{\alpha}_i \geq 0, \quad (f(x_i) - 1) y_i \geq 0$

Large Margin SVM

Primal Problem:

$$\text{Minimize } \tfrac{1}{2}\|w\|^2 \text{ subject to } \forall i \in [n]: \ y_i(w^T x_i + w_0) \geq 1, \quad w \in \mathbb{R}^p,\ w_0 \in \mathbb{R}$$

Lagrangian: $L(w, w_0, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^n \alpha_i \left(y_i(w^T x_i + w_0) - 1 \right), \quad w \in \mathbb{R}^p, w_0 \in \mathbb{R}, \alpha \in \mathbb{R}_+^p$

Optimization theory Theorem: There exists $\hat{w}, \hat{w}_0, \hat{\alpha}$ s.t. (denoting $f : x \mapsto \hat{w}^T x + \hat{w}_0$):

- $\min\limits_{w \in \mathbb{R}^p, w_0 \in \mathbb{R}} \max\limits_{\alpha \in \mathbb{R}_+^n} L(w, w_0, \alpha) = L(\hat{w}, \hat{w}_0, \hat{\alpha}) = \max\limits_{\alpha \in \mathbb{R}_+^n} \min\limits_{w \in \mathbb{R}^p, w_0 \in \mathbb{R}} L(w, w_0, \alpha)$
- KKT conditions: $\frac{\partial L}{\partial w}\big|_{\hat{w}, \hat{w}_0, \hat{\alpha}} = 0, \qquad \frac{\partial L}{\partial w_0}\big|_{\hat{w}, \hat{w}_0, \hat{\alpha}} = 0,$

  $\forall i \in [n]: \qquad \hat{\alpha}_i(f(x_i) - 1)y_i = 0, \qquad \hat{\alpha}_i \geq 0, \qquad (f(x_i) - 1)y_i \geq 0$

Large Margin SVM

Primal Problem:

$$\text{Minimize } \tfrac{1}{2}\|w\|^2 \text{ subject to } \forall i \in [n]: \ y_i(w^T x_i + w_0) \geq 1, \quad w \in \mathbb{R}^p,\ w_0 \in \mathbb{R}$$

Lagrangian: $L(w, w_0, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^n \alpha_i \left(y_i(w^T x_i + w_0) - 1 \right), \quad w \in \mathbb{R}^p, w_0 \in \mathbb{R}, \alpha \in \mathbb{R}_+^p$

Optimization theory Theorem: There exists $\hat{w}, \hat{w}_0, \hat{\alpha}$ s.t. (denoting $f : x \mapsto \hat{w}^T x + \hat{w}_0$):

- $\min_{w \in \mathbb{R}^p, w_0 \in \mathbb{R}} \max_{\alpha \in \mathbb{R}_+^n} L(w, w_0, \alpha) = L(\hat{w}, \hat{w}_0, \hat{\alpha}) = \max_{\alpha \in \mathbb{R}_+^n} \min_{w \in \mathbb{R}^p, w_0 \in \mathbb{R}} L(w, w_0, \alpha)$
- KKT conditions: $\frac{\partial L}{\partial w}\Big|_{\hat{w}, \hat{w}_0, \hat{\alpha}} = 0, \qquad \frac{\partial L}{\partial w_0}\Big|_{\hat{w}, \hat{w}_0, \hat{\alpha}} = 0,$

$$\forall i \in [n]: \qquad \hat{\alpha}_i (f(x_i) - 1) y_i = 0, \qquad \hat{\alpha}_i \geq 0, \qquad (f(x_i) - 1) y_i \geq 0$$

Resolution:

- $\hat{w} = \sum_{i=1}^n \hat{\alpha}_i y_i x_i$

Large Margin SVM

Primal Problem:

$$\text{Minimize } \tfrac{1}{2}\|w\|^2 \text{ subject to } \forall i \in [n] :\ \ y_i(w^T x_i + w_0) \geq 1, \quad w \in \mathbb{R}^p,\ w_0 \in \mathbb{R}$$

Lagrangian: $L(w, w_0, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^n \alpha_i \left(y_i(w^T x_i + w_0) - 1\right), \quad w \in \mathbb{R}^p, w_0 \in \mathbb{R}, \alpha \in \mathbb{R}^p_+$

Optimization theory Theorem: There exists $\hat{w}, \hat{w}_0, \hat{\alpha}$ s.t. (denoting $f : x \mapsto \hat{w}^T x + \hat{w}_0$):

- $\min_{w \in \mathbb{R}^p, w_0 \in \mathbb{R}} \max_{\alpha \in \mathbb{R}^n_+} L(w, w_0, \alpha) = L(\hat{w}, \hat{w}_0, \hat{\alpha}) = \max_{\alpha \in \mathbb{R}^n_+} \min_{w \in \mathbb{R}^p, w_0 \in \mathbb{R}} L(w, w_0, \alpha)$
- KKT conditions: $\frac{\partial L}{\partial w}\big|_{\hat{w}, \hat{w}_0, \hat{\alpha}} = 0, \qquad \frac{\partial L}{\partial w_0}\big|_{\hat{w}, \hat{w}_0, \hat{\alpha}} = 0,$

  $\forall i \in [n] : \qquad \hat{\alpha}_i (f(x_i) - 1) y_i = 0, \qquad \hat{\alpha}_i \geq 0, \qquad (f(x_i) - 1) y_i \geq 0$

Resolution:

- $\hat{w} = \sum_{i=1}^n \hat{\alpha}_i y_i x_i$
- $\sum_{i=1}^n \hat{\alpha}_i y_i = 0$

Large Margin SVM

Primal Problem:

$$\text{Minimize } \tfrac{1}{2}\|w\|^2 \text{ subject to } \forall i \in [n]: \; y_i(w^T x_i + w_0) \geq 1, \quad w \in \mathbb{R}^p,\, w_0 \in \mathbb{R}$$

Lagrangian: $L(w, w_0, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^n \alpha_i \left(y_i(w^T x_i + w_0) - 1\right), \quad w \in \mathbb{R}^p, w_0 \in \mathbb{R}, \alpha \in \mathbb{R}_+^p$

Optimization theory Theorem: There exists $\hat{w}, \hat{w}_0, \hat{\alpha}$ s.t. (denoting $f : x \mapsto \hat{w}^T x + \hat{w}_0$):

- $\min\limits_{w\in\mathbb{R}^p, w_0\in\mathbb{R}} \max\limits_{\alpha\in\mathbb{R}_+^n} L(w, w_0, \alpha) = L(\hat{w}, \hat{w}_0, \hat{\alpha}) = \max\limits_{\alpha\in\mathbb{R}_+^n} \min\limits_{w\in\mathbb{R}^p, w_0\in\mathbb{R}} L(w, w_0, \alpha)$
- KKT conditions: $\frac{\partial L}{\partial w}\Big|_{\hat{w},\hat{w}_0,\hat{\alpha}} = 0, \qquad \frac{\partial L}{\partial w_0}\Big|_{\hat{w},\hat{w}_0,\hat{\alpha}} = 0,$

$$\forall i \in [n]: \qquad \hat{\alpha}_i(f(x_i) - 1)y_i = 0, \qquad \hat{\alpha}_i \geq 0, \qquad (f(x_i) - 1)y_i \geq 0$$

Resolution:

- $\hat{w} = \sum_{i=1}^n \hat{\alpha}_i y_i x_i$
- $\sum_{i=1}^n \hat{\alpha}_i y_i = 0$

Dual Problem: Maximize $-\frac{1}{2}\sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j x_i x_j + \sum_{i=1}^n \alpha_i, \quad \alpha \in \mathbb{R}_+^n$

Large Margin SVM

Primal Problem:

$$\text{Minimize } \tfrac{1}{2}\|w\|^2 \text{ subject to } \forall i \in [n] : \; y_i(w^T x_i + w_0) \geq 1, \quad w \in \mathbb{R}^p, \; w_0 \in \mathbb{R}$$

Lagrangian: $L(w, w_0, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^n \alpha_i \left(y_i(w^T x_i + w_0) - 1 \right), \quad w \in \mathbb{R}^p, w_0 \in \mathbb{R}, \alpha \in \mathbb{R}_+^p$

Optimization theory Theorem: There exists $\hat{w}, \hat{w}_0, \hat{\alpha}$ s.t. (denoting $f : x \mapsto \hat{w}^T x + \hat{w}_0$):

- $\min_{w \in \mathbb{R}^p, w_0 \in \mathbb{R}} \max_{\alpha \in \mathbb{R}_+^n} L(w, w_0, \alpha) = L(\hat{w}, \hat{w}_0, \hat{\alpha}) = \max_{\alpha \in \mathbb{R}_+^n} \min_{w \in \mathbb{R}^p, w_0 \in \mathbb{R}} L(w, w_0, \alpha)$
- KKT conditions: $\frac{\partial L}{\partial w}\big|_{\hat{w}, \hat{w}_0, \hat{\alpha}} = 0, \qquad \frac{\partial L}{\partial w_0}\big|_{\hat{w}, \hat{w}_0, \hat{\alpha}} = 0,$

  $\forall i \in [n] : \qquad \hat{\alpha}_i (f(x_i) - 1) y_i = 0, \qquad \hat{\alpha}_i \geq 0, \qquad (f(x_i) - 1) y_i \geq 0$

Resolution:

- $\hat{w} = \sum_{i=1}^n \hat{\alpha}_i y_i x_i$
- $\sum_{i=1}^n \hat{\alpha}_i y_i = 0$

Dual Problem: Maximize $-\frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j x_i x_j + \sum_{i=1}^n \alpha_i, \quad \alpha \in \mathbb{R}_+^n$

Solution $\hat{w}, \hat{w}_0, \hat{\alpha}$ satisfy:

$\forall i \in [n] : \alpha_i > 0 \implies f(x_i) y_i = 1$: Support point

Large Margin SVM

Primal Problem:

$$\text{Minimize } \frac{1}{2}\|w\|^2 \text{ subject to } \forall i \in [n]: \ y_i(w^T x_i + w_0) \geq 1, \quad w \in \mathbb{R}^p,\ w_0 \in \mathbb{R}$$

Lagrangian: $L(w, w_0, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^n \alpha_i \left(y_i(w^T x_i + w_0) - 1\right), \quad w \in \mathbb{R}^p, w_0 \in \mathbb{R}, \alpha \in \mathbb{R}_+^p$

Resolution:

- $\hat{w} = \sum_{i=1}^n \hat{\alpha}_i y_i x_i$
- $\sum_{i=1}^n \hat{\alpha}_i y_i = 0$

Dual Problem: Maximize $-\frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j x_i x_j + \sum_{i=1}^n \alpha_i, \quad \alpha \in \mathbb{R}_+^n$

Solution $\hat{w}, \hat{w}_0, \hat{\alpha}$ satisfy:

$\forall i \in [n] : \alpha_i > 0 \implies (\hat{w}^T x_i - \hat{w}_0) y_i = 1$: Support points

Large Margin SVM

Primal Problem:

$$\text{Minimize } \tfrac{1}{2}\|w\|^2 \text{ subject to } \forall i \in [n]: \; y_i(w^T x_i + w_0) \geq 1, \quad w \in \mathbb{R}^p,\ w_0 \in \mathbb{R}$$

Lagrangian: $L(w, w_0, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^n \alpha_i \left(y_i(w^T x_i + w_0) - 1\right), \quad w \in \mathbb{R}^p, w_0 \in \mathbb{R}, \alpha \in \mathbb{R}_+^p$

Resolution:

- $\hat{w} = \sum_{i=1}^n \hat{\alpha}_i y_i x_i$
- $\sum_{i=1}^n \hat{\alpha}_i y_i = 0$

Dual Problem: Maximize $-\frac{1}{2}\sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j x_i x_j + \sum_{i=1}^n \alpha_i, \quad \alpha \in \mathbb{R}_+^n$

Solution $\hat{w}, \hat{w}_0, \hat{\alpha}$ satisfy:

$\forall i \in [n] : \alpha_i > 0 \implies (\hat{w}^T x_i - \hat{w}_0) y_i = 1$: Support points

Final prediction: $f(x) = \sum_{i=1}^n \hat{\alpha}_i y_i x_i^T x + \hat{w}_0$

Large Margin SVM

Primal Problem:

$$\text{Minimize } \tfrac{1}{2}\|w\|^2 \text{ subject to } \forall i \in [n] : \ y_i(w^T x_i + w_0) \geq 1, \quad w \in \mathbb{R}^p, \ w_0 \in \mathbb{R}$$

Lagrangian: $L(w, w_0, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^n \alpha_i \left(y_i(w^T x_i + w_0) - 1 \right), \quad w \in \mathbb{R}^p, w_0 \in \mathbb{R}, \alpha \in \mathbb{R}_+^p$

Resolution:

- $\hat{w} = \sum_{i=1}^n \hat{\alpha}_i y_i x_i$
- $\sum_{i=1}^n \hat{\alpha}_i y_i = 0$

Dual Problem: Maximize $-\frac{1}{2}\sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j x_i x_j + \sum_{i=1}^n \alpha_i, \quad \alpha \in \mathbb{R}_+^n$

Solution $\hat{w}, \hat{w}_0, \hat{\alpha}$ satisfy:

$$\forall i \in [n] : \alpha_i > 0 \implies (\hat{w}^T x_i - \hat{w}_0) y_i = 1 : \{\text{Support points}\} \equiv S$$

Final prediction: $f(x) = \sum_{i \in S} \hat{\alpha}_i y_i x_i^T x + \hat{w}_0$

Large Margin SVM

Primal Problem:

$$\text{Minimize } \tfrac{1}{2}\|w\|^2 \text{ subject to } \forall i \in [n]: \; y_i(w^T x_i + w_0) \geq 1, \quad w \in \mathbb{R}^p,\, w_0 \in \mathbb{R}$$

Lagrangian: $L(w, w_0, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^n \alpha_i \left(y_i(w^T x_i + w_0) - 1\right), \quad w \in \mathbb{R}^p, w_0 \in \mathbb{R}, \alpha \in \mathbb{R}_+^p$

Resolution:

- $\hat{w} = \sum_{i=1}^n \hat{\alpha}_i y_i x_i$
- $\sum_{i=1}^n \hat{\alpha}_i y_i = 0$

Dual Problem: Maximize $-\frac{1}{2}\sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j x_i x_j + \sum_{i=1}^n \alpha_i, \quad \alpha \in \mathbb{R}_+^n$

Solution $\hat{w}, \hat{w}_0, \hat{\alpha}$ satisfy:

$$\forall i \in [n] : \alpha_i > 0 \implies (\hat{w}^T x_i - \hat{w}_0) y_i = 1 : \{\text{Support points}\} \equiv S$$

Final prediction: $f(x) = \sum_{i \in S} \hat{\alpha}_i y_i x_i^T x + \hat{w}_0$

How to compute $\hat{w}_0$?

Large Margin SVM

Primal Problem:

$$\text{Minimize } \tfrac{1}{2}\|w\|^2 \text{ subject to } \forall i \in [n]: \ y_i(w^T x_i + w_0) \geq 1, \quad w \in \mathbb{R}^p,\ w_0 \in \mathbb{R}$$

Lagrangian: $L(w, w_0, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^n \alpha_i \left(y_i(w^T x_i + w_0) - 1\right), \quad w \in \mathbb{R}^p, w_0 \in \mathbb{R}, \alpha \in \mathbb{R}_+^p$

Resolution:

- $\hat{w} = \sum_{i=1}^n \hat{\alpha}_i y_i x_i$
- $\sum_{i=1}^n \hat{\alpha}_i y_i = 0$

Dual Problem: Maximize $-\frac{1}{2}\sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j x_i x_j + \sum_{i=1}^n \alpha_i, \quad \alpha \in \mathbb{R}_+^n$

Solution $\hat{w}, \hat{w}_0, \hat{\alpha}$ satisfy:

$$\forall i \in [n] : \alpha_i > 0 \implies (\hat{w}^T x_i - \hat{w}_0) y_i = 1 : \{\text{Support points}\} \equiv S$$

Final prediction: $f(x) = \sum_{i \in S} \hat{\alpha}_i y_i x_i^T x + \hat{w}_0$

$$\forall j \in S : 1 = f(x_j) y_j = \sum_{i \in S} \hat{\alpha}_i y_i y_j x_i^T x_j + \hat{w}_0 y_j$$

Large Margin SVM

Primal Problem:

$$\text{Minimize } \tfrac{1}{2}\|w\|^2 \text{ subject to } \forall i \in [n]: \; y_i(w^T x_i + w_0) \geq 1, \quad w \in \mathbb{R}^p,\ w_0 \in \mathbb{R}$$

Lagrangian: $L(w, w_0, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^n \alpha_i \left(y_i(w^T x_i + w_0) - 1 \right), \quad w \in \mathbb{R}^p, w_0 \in \mathbb{R}, \alpha \in \mathbb{R}_+^p$

Resolution:

- $\hat{w} = \sum_{i=1}^n \hat{\alpha}_i y_i x_i$
- $\sum_{i=1}^n \hat{\alpha}_i y_i = 0$

Dual Problem: Maximize $-\frac{1}{2}\sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j x_i x_j + \sum_{i=1}^n \alpha_i, \quad \alpha \in \mathbb{R}_+^n$

Solution $\hat{w}, \hat{w}_0, \hat{\alpha}$ satisfy:

$$\forall i \in [n] : \alpha_i > 0 \implies (\hat{w}^T x_i - \hat{w}_0) y_i = 1 : \{\text{Support points}\} \equiv S$$

Final prediction: $f(x) = \sum_{i \in S} \hat{\alpha}_i y_i x_i^T x + \hat{w}_0$

$y_j^2 = 1$

$$\forall j \in S : 1 = f(x_j) y_j = \sum_{i \in S} \hat{\alpha}_i y_i y_j x_i^T x_j + \hat{w}_0 y_j \implies \hat{w}_0 = y_j - \sum_{i \in S} \hat{\alpha}_i y_i x_i^T x_j$$

Large Margin SVM

Primal Problem:

$$\text{Minimize } \tfrac{1}{2}\|w\|^2 \text{ subject to } \forall i \in [n]:\ y_i(w^T x_i + w_0) \geq 1, \quad w \in \mathbb{R}^p,\ w_0 \in \mathbb{R}$$

Lagrangian: $L(w, w_0, \alpha) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^n \alpha_i \left(y_i(w^T x_i + w_0) - 1\right), \quad w \in \mathbb{R}^p, w_0 \in \mathbb{R}, \alpha \in \mathbb{R}_+^p$

Resolution:

- $\hat{w} = \sum_{i=1}^n \hat{\alpha}_i y_i x_i$
- $\sum_{i=1}^n \hat{\alpha}_i y_i = 0$

Dual Problem: Maximize $-\frac{1}{2}\sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j x_i x_j + \sum_{i=1}^n \alpha_i, \quad \alpha \in \mathbb{R}_+^n$

Solution $\hat{w}, \hat{w}_0, \hat{\alpha}$ satisfy:

$$\forall i \in [n]: \alpha_i > 0 \implies (\hat{w}^T x_i - \hat{w}_0) y_i = 1 \text{:\{Support points\}} \equiv S$$

Final prediction: $f(x) = \sum_{i \in S} \hat{\alpha}_i y_i x_i^T x + \hat{w}_0$

$$\forall j \in S: 1 = f(x_j) y_j = \sum_{i \in S} \hat{\alpha}_i y_i y_j x_i^T x_j + \hat{w}_0 y_j \qquad \implies \hat{w}_0 = y_j - \sum_{i \in S} \hat{\alpha}_i y_i x_i^T x_j$$

In practice averaging: $\hat{w}_0 \equiv \frac{1}{|S|} \sum_{j \in S} y_j - \sum_{i \in S} \hat{\alpha}_i y_i x_i^T x_j$

Large Margin SVM

Primal Problem:

Minimize $\frac{1}{2}\|w\|^2$ subject to $\forall i \in [n]: \; y_i(w^T x_i + w_0) \geq 1$

Dual Problem:

Maximize $-\frac{1}{2}\sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i x_j + \sum_{i=1}^{n} \alpha_i$

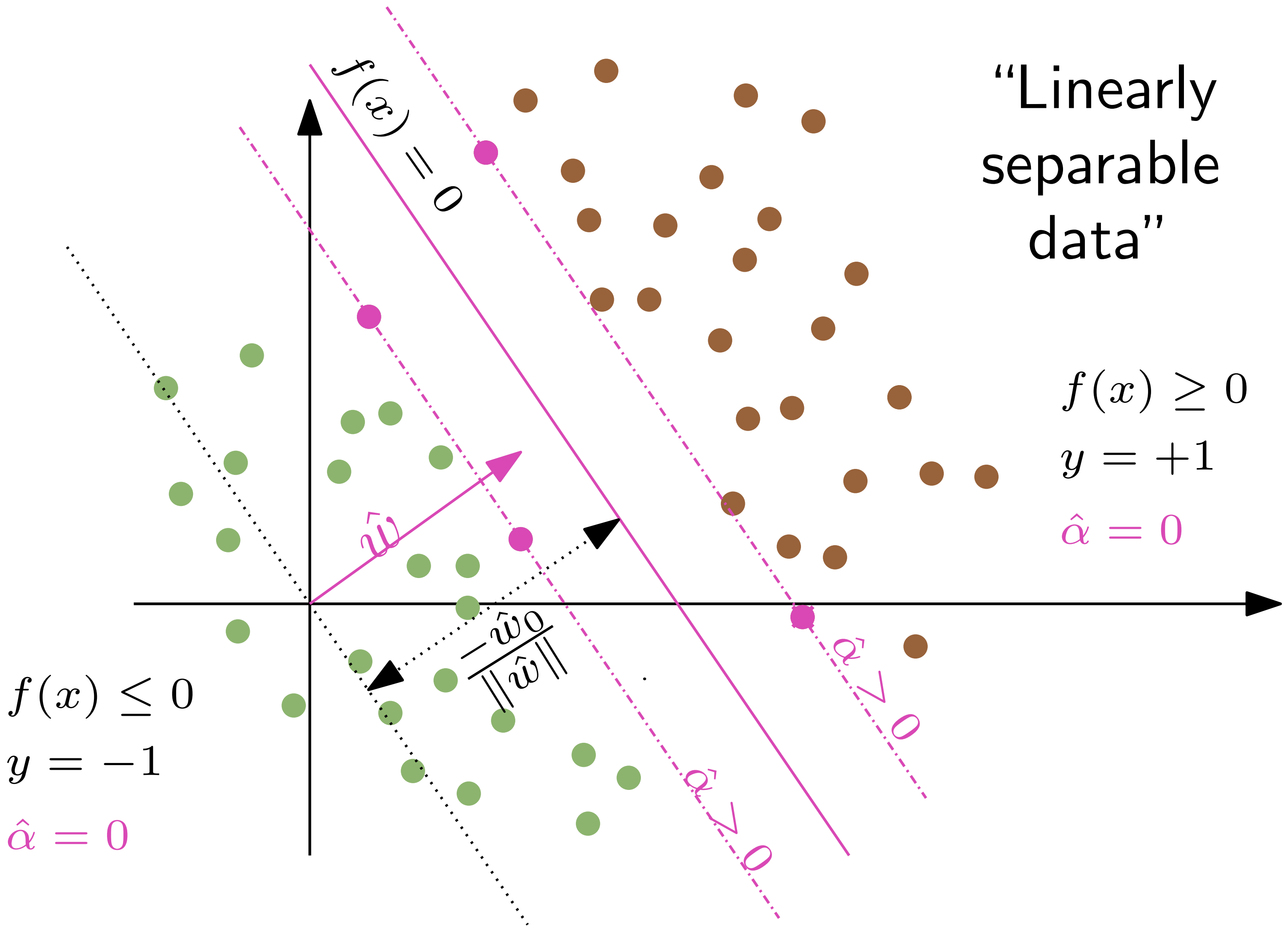
Subject to $\alpha \in \mathbb{R}_+^n$, $\sum_{i=1}^{n} \alpha_i y_i = 0$.

Final prediction:

$$f(x) = \hat{w}^T x + \hat{w}_0$$

with: $\hat{w} = \sum_{i \in S} \hat{\alpha}_i y_i x_i \qquad S = \{i : \alpha_i > 0\}$

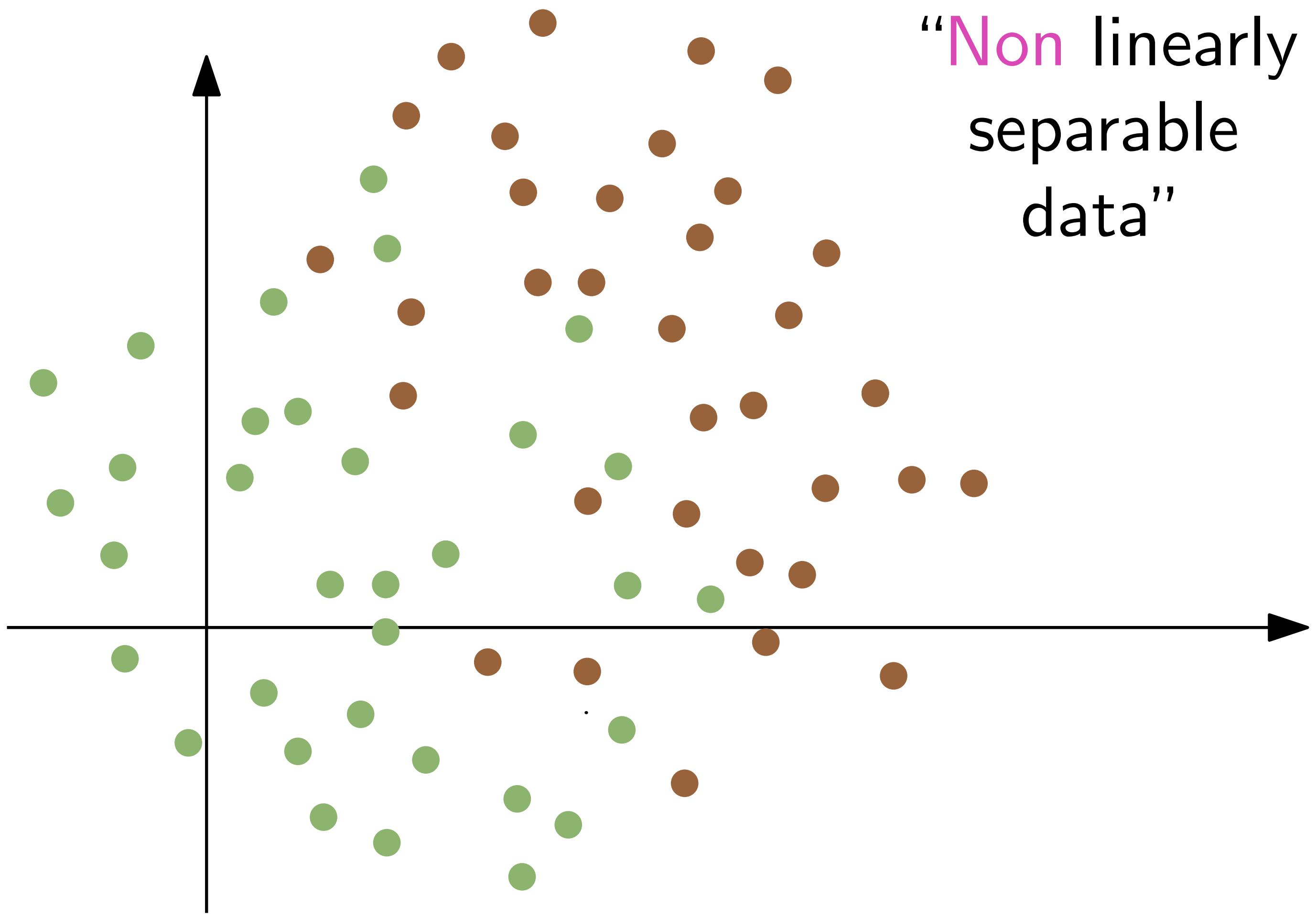
$$\hat{w}_0 = \frac{1}{|S|} \sum_{j \in S} y_j - \sum_{i \in S} \hat{\alpha}_i y_i x_i^T x_j$$



Soft Margin SVM

Primal Problem:

Minimize $\frac{1}{2}\|w\|^2$ subject to $\forall i \in [n]: \; y_i(w^T x_i + w_0) \geq 1$

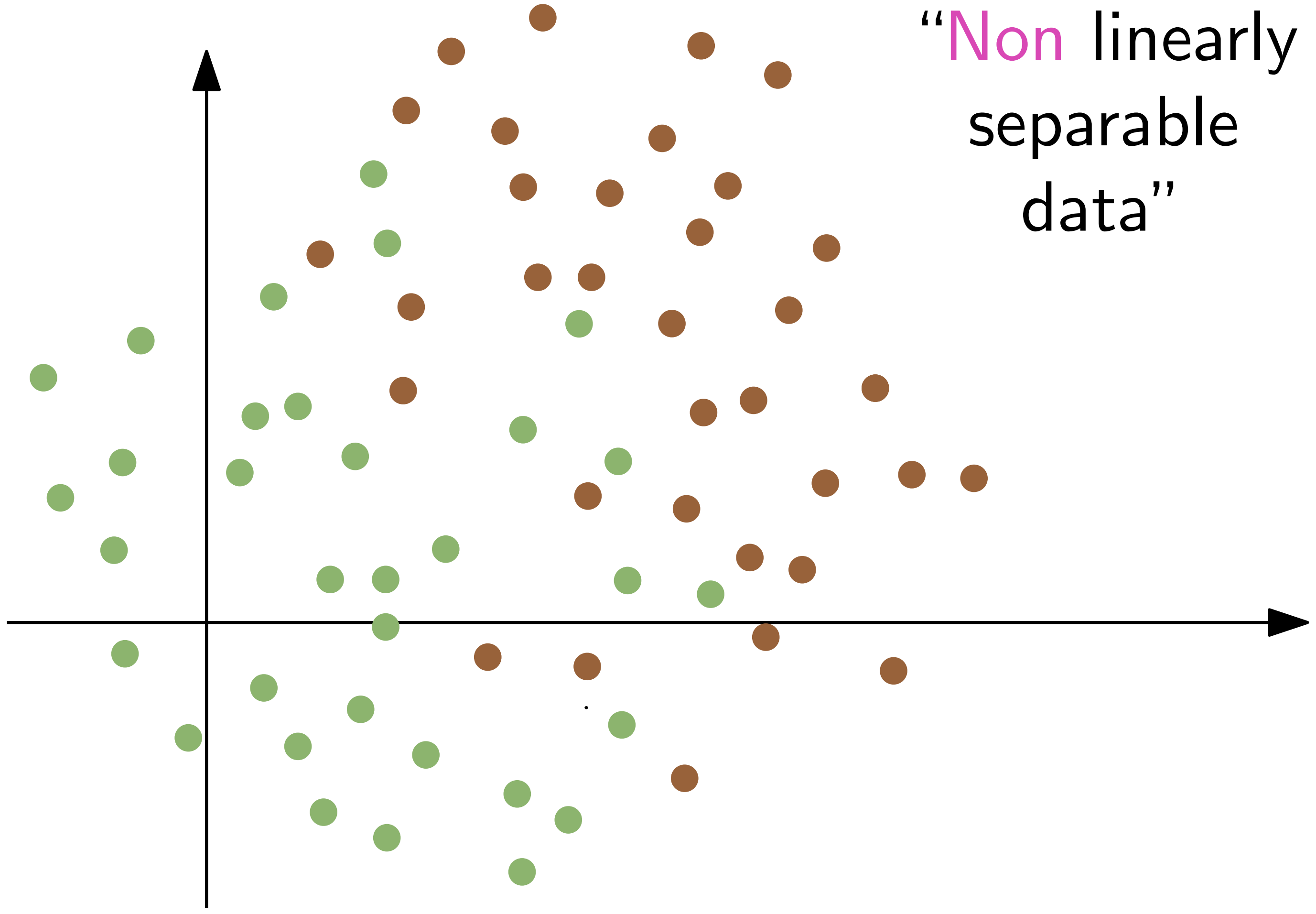


"Non linearly separable data"

Soft Margin SVM

Primal Problem:

Minimize $\frac{1}{2}\|w\|^2 + C\sum_{i=1}^n \xi_i$ subject to $\forall i \in [n]: \;\; y_i(w^T x_i + w_0) \geq 1 - \xi_i, \quad \xi_i > 0$



"Non linearly separable data"

Soft Margin SVM

Primal Problem:

Minimize $\frac{1}{2}\|w\|^2 + C\sum_{i=1}^n \xi_i$ subject to $\forall i \in [n]: \; y_i(w^T x_i + w_0) \geq 1 - \xi_i, \quad \xi_i > 0$

Dual Problem:

Maximize $-\frac{1}{2}\sum_{i,j=1}^n \alpha_i\alpha_j y_i y_j x_i x_j + \sum_{i=1}^n \alpha_i$

Subject to $\alpha \in [0, C]^n$, $\sum_{i=1}^n \alpha_i y_i = 0$.

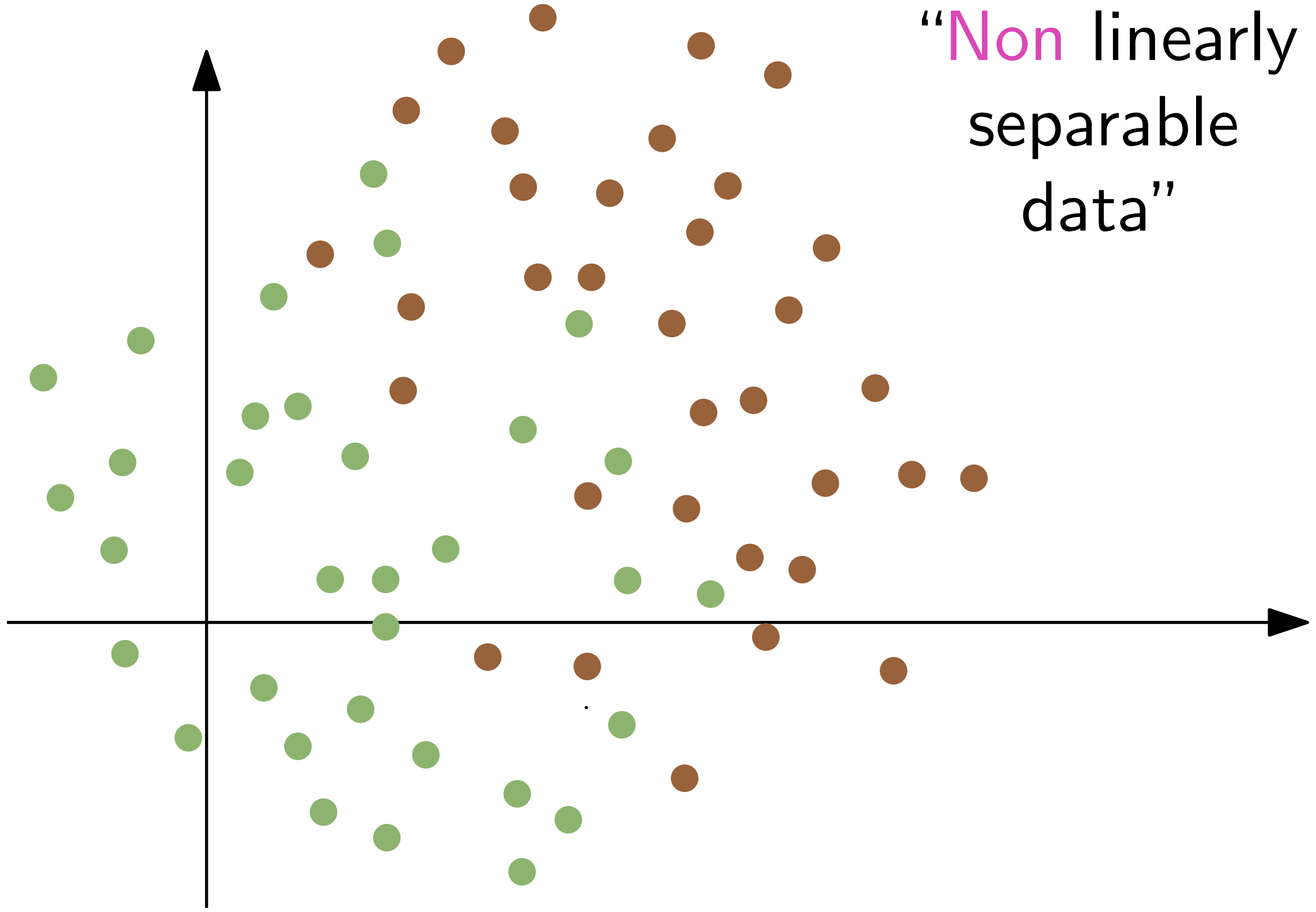
Final prediction:

$$f(x) = \hat{w}^T x + \hat{w}_0$$

(Same kind of inferences as before)

with: $$\hat{w} = \sum_{i \in S} \hat{\alpha}_i y_i x_i$$

$$\hat{w}_0 = \frac{1}{|S|}\sum_{j \in S} y_j - \sum_{i \in S} \hat{\alpha}_i y_i x_i^T x_j$$



"Non linearly separable data"

Soft Margin SVM

Primal Problem:

Minimize $\frac{1}{2}\|w\|^2 + C\sum_{i=1}^n \xi_i$ subject to $\forall i \in [n]: \ y_i(w^T x_i + w_0) \geq 1 - \xi_i, \quad \xi_i > 0$

Dual Problem:

Maximize $-\frac{1}{2}\sum_{i,j=1}^n \alpha_i\alpha_j y_i y_j x_i x_j + \sum_{i=1}^n \alpha_i$

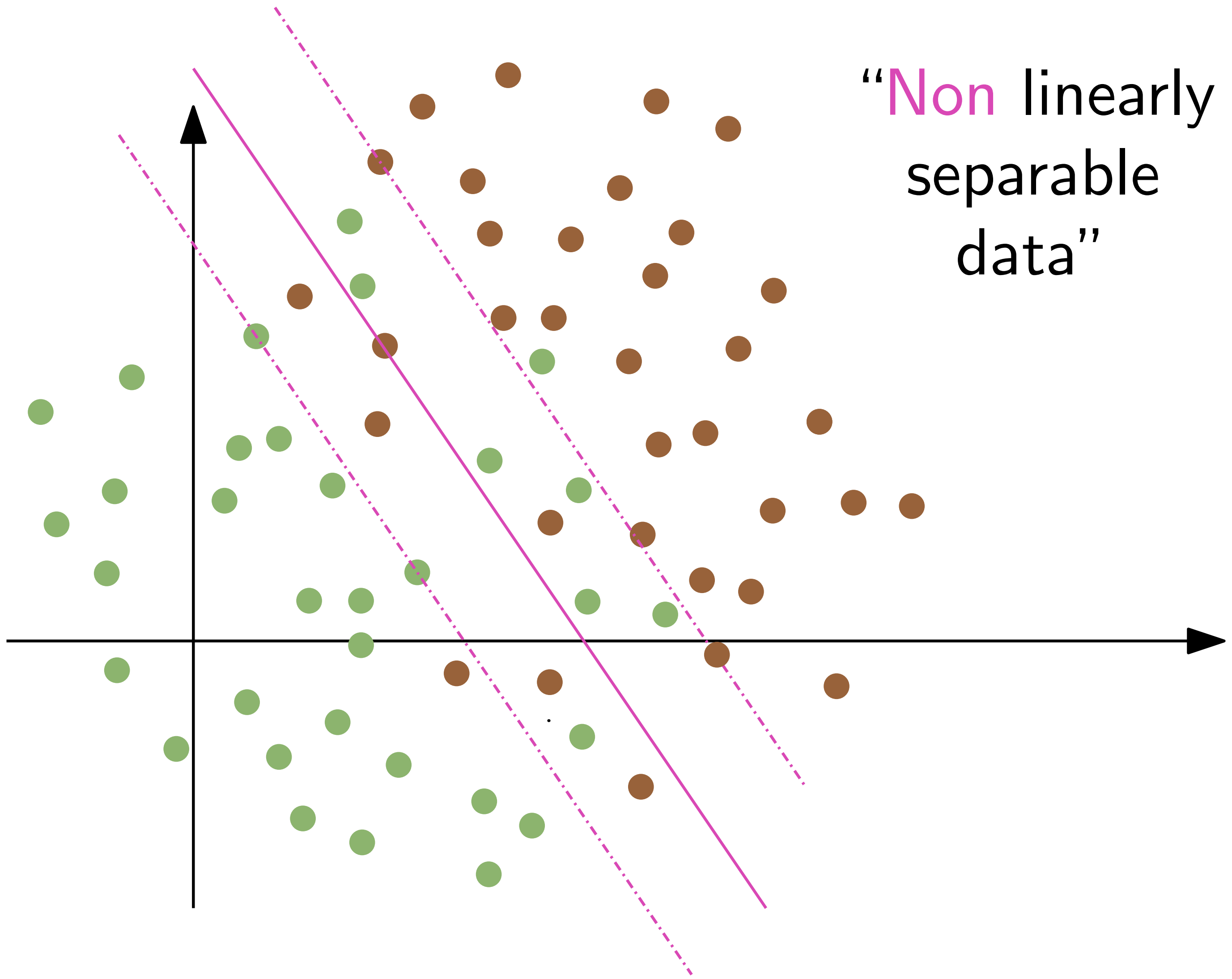
Subject to $\alpha \in [0, C]^n$, $\sum_{i=1}^n \alpha_i y_i = 0$.

Final prediction:

$$f(x) = \hat{w}^T x + \hat{w}_0$$

with: $$\hat{w} = \sum_{i \in S} \hat{\alpha}_i y_i x_i$$

$$\hat{w}_0 = \frac{1}{|S|}\sum_{j \in S} y_j - \sum_{i \in S} \hat{\alpha}_i y_i x_i^T x_j$$



"Non linearly separable data"

Soft Margin SVM

Primal Problem:

Minimize $\frac{1}{2}\|w\|^2 + C\sum_{i=1}^n \xi_i$ subject to $\forall i \in [n]:\ y_i(w^T x_i + w_0) \geq 1 - \xi_i, \quad \xi_i > 0$

Dual Problem:

Maximize $-\frac{1}{2}\sum_{i,j=1}^n \alpha_i\alpha_j y_i y_j x_i x_j + \sum_{i=1}^n \alpha_i$

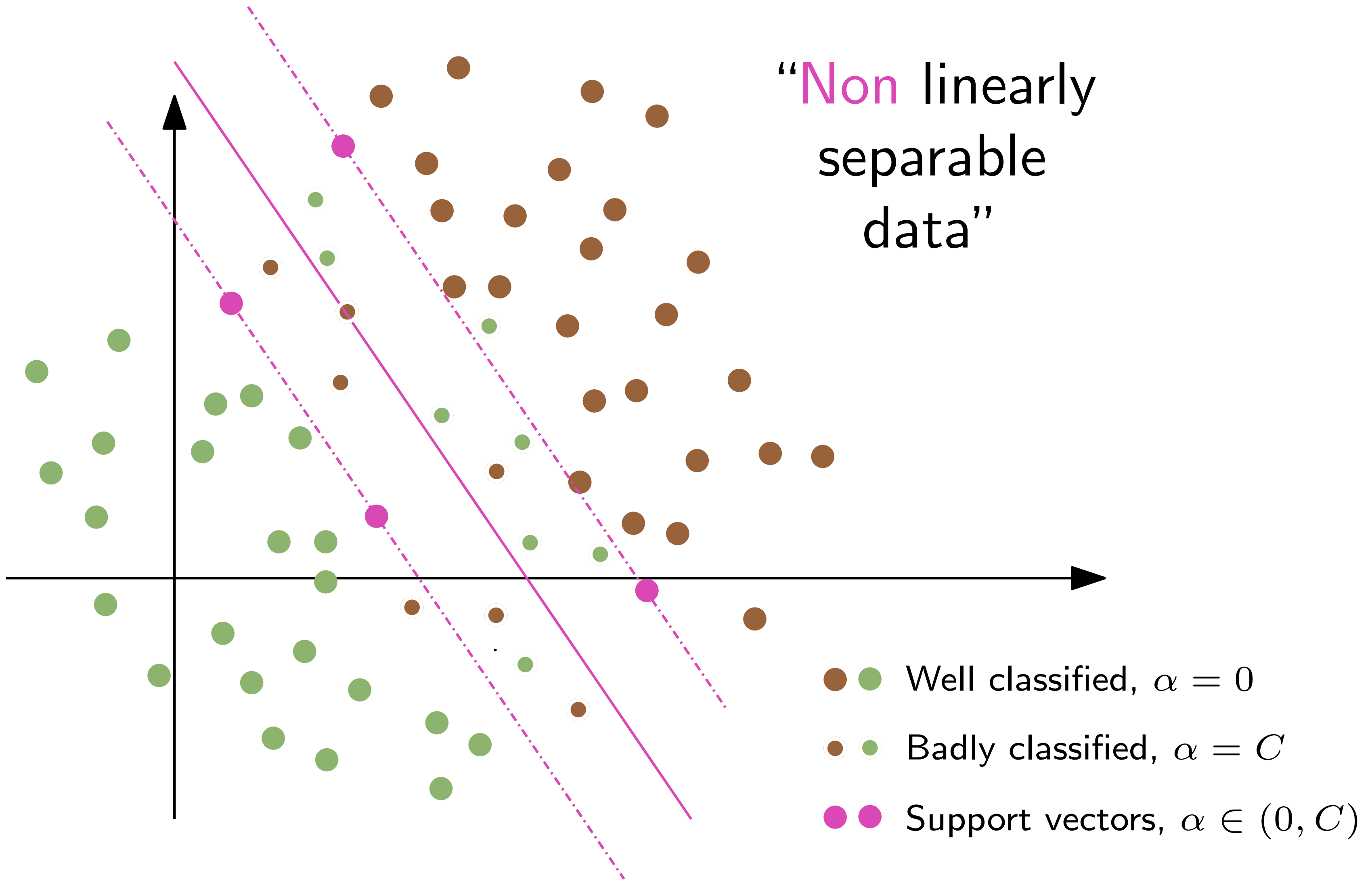
Subject to $\alpha \in [0, C]^n$, $\sum_{i=1}^n \alpha_i y_i = 0$.

Final prediction:

$$f(x) = \hat{w}^T x + \hat{w}_0$$

with: $\hat{w} = \sum_{i \in S} \hat{\alpha}_i y_i x_i$

$$\hat{w}_0 = \frac{1}{|S|}\sum_{j \in S} y_j - \sum_{i \in S} \hat{\alpha}_i y_i x_i^T x_j$$



Kernel trick with SVM

Minimize: $\sum_{i,j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j - \sum_{i=1}^{n} \alpha_i$

Subject to:
- $\forall i \in [n] : 0 \leq \alpha_i \leq C$
- $\sum_{i=1}^{n} \alpha_i y_i = 0$

Then:
- $f(t) = \sum_{i \in S} \hat{\alpha}_i y_i x^T x_i + \hat{w}_0$
- $\hat{w}_0 = \frac{1}{|S|} \sum_{i \in S} \left(y_i - \sum_{j \in S} \hat{\alpha}_j y_j x_j^T x_i \right)$

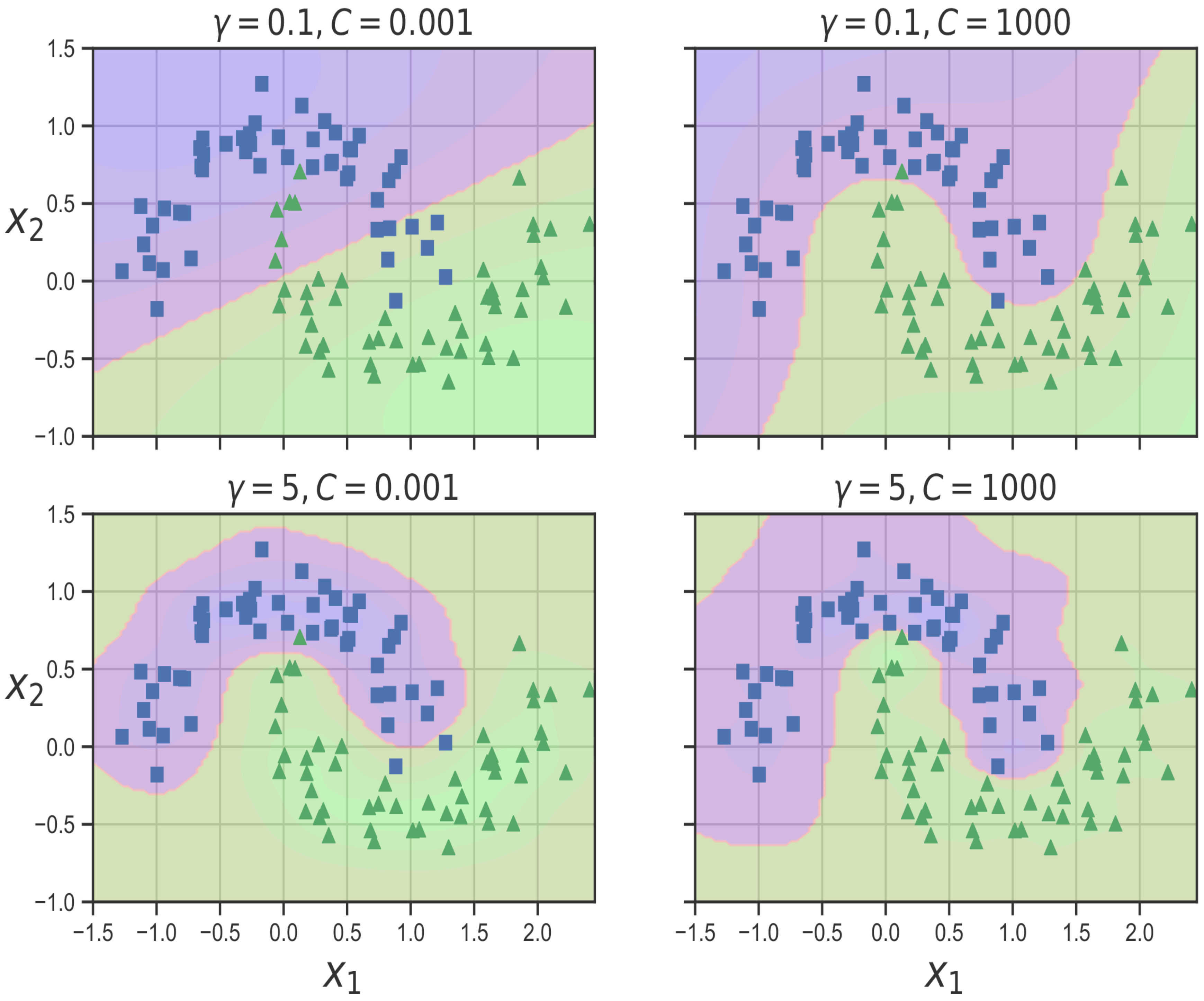
Kernel trick with SVM

Minimize: $\sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^n \alpha_i$

Subject to:
- $\forall i \in [n] : 0 \leq \alpha_i \leq C$
- $\displaystyle\sum_{i=1}^n \alpha_i y_i = 0$

Then:
- $\displaystyle f(t) = \sum_{i \in S} \hat{\alpha}_i y_i K(x, x_i) + \hat{w}_0$
- $\displaystyle \hat{w}_0 = \frac{1}{|S|} \sum_{i \in S} \left(y_i - \sum_{j \in S} \hat{\alpha}_j y_j K(x_j, x_i) \right)$

Kernel trick with SVM



Minimize: $\sum_{i,j=1}^n \alpha_i\alpha_j y_i y_j K(x_i, x_j) - \sum_{i=1}^n \alpha_i$

Subject to:
- $\forall i \in [n] : 0 \leq \alpha_i \leq C$
- $\displaystyle\sum_{i=1}^n \alpha_i y_i = 0$

Then:
- $\displaystyle f(t) = \sum_{i\in S} \hat\alpha_i y_i K(x, x_i) + \hat w_0$
- $\displaystyle \hat w_0 = \frac{1}{|S|}\sum_{i\in S}\left(y_i - \sum_{j\in S}\hat\alpha_j y_j K(x_j, x_i)\right)$

Exmple with $K(x,z) = \exp\left(-\gamma\|x-z\|^2\right)$