Machine Learning I
60629A

Summary
Supervised Learning
— Week #3

Today: Models for supervised learning

- (Mostly) linear models
- Focus on classification

1. Non-Probabilistic Models
   - Nearest Neighbor (k-NN), Support Vector Machines (SVMs)
2. Probabilistic Models
   - Naive Bayes

Supervised learning

Train Data

	Nb.bed.	Area	Neigh.	·	·		Sell-ability
x_0	1	0	0	0	0	y_0	1
x_1	1	100	1	.2	.5	y_1	2
x_2	3	200	0	.1	.2	y_2	0
x_3	1	150	1	.4	.1	y_3	2
x_4	2	210	2	.5	1.1	y_4	1
	X						Y

Task

$$f : \mathbb{R}^n \rightarrow \{0, 1, 2\}$$

Test Data

	Nb.bed.	Area	Neigh.	·	·		Sell-ability
x_0	1	0	0	0	0	y_0	?
x_1	2	50	1	.3	.8	y_1	?
x_2	1	100	1	.5	1.4	y_2	?
x_3	4	170	0	.7	.4	y_3	?
x_4	1	120	3	.9	.5	y_4	?
	X^{new}						Y^{new}

Supervised learning

Train Data

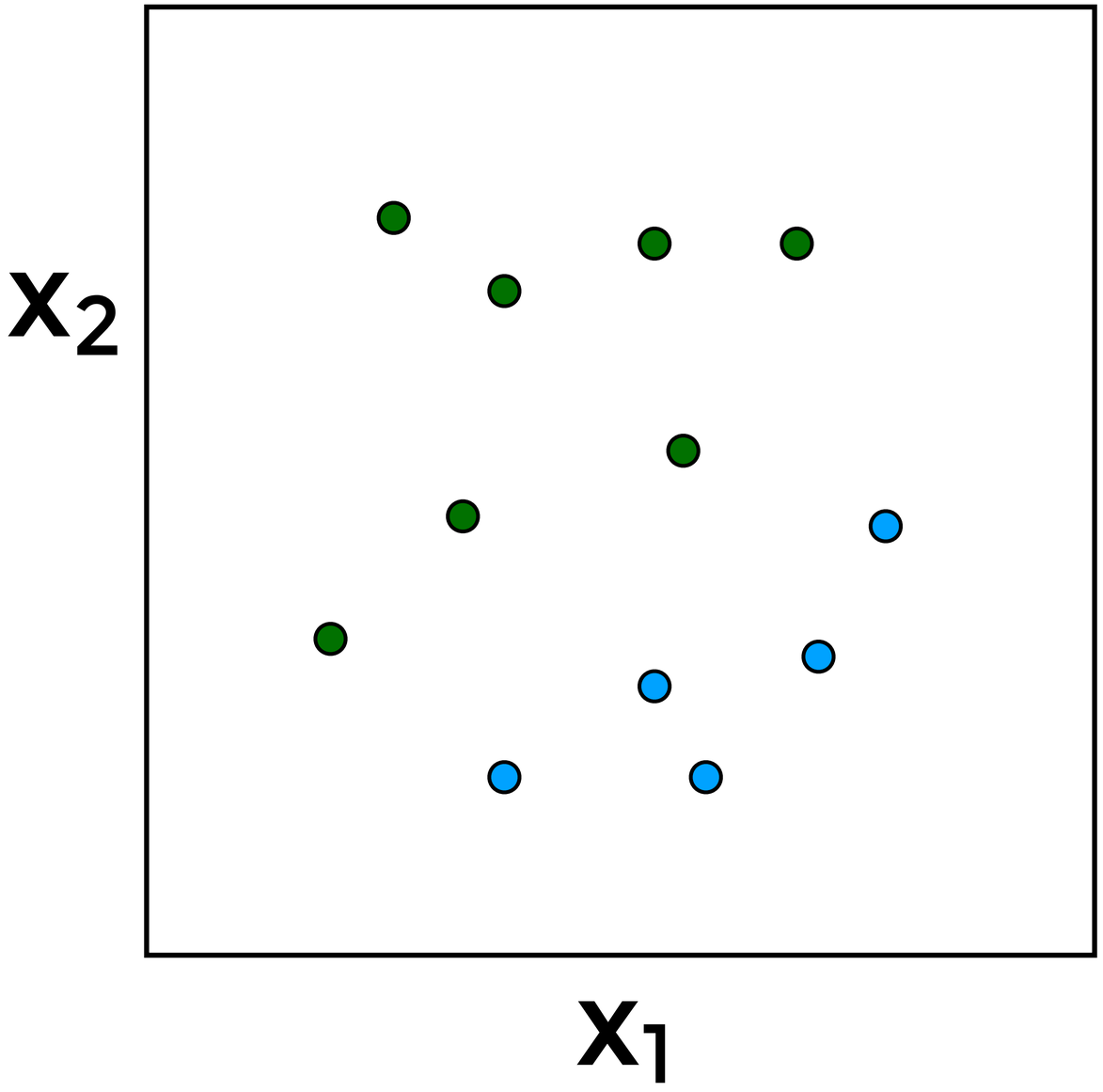
	Nb.bed.	Area	Neigh.	·	·		Sell-ability
x_0	1	0	0	0	0	y_0	1
x_1	1	100	1	.2	.5	y_1	2
x_2	3	200	0	.1	.2	y_2	0
x_3	1	150	1	.4	.1	y_3	2
x_4	2	210	2	.5	1.1	y_4	1
	X						Y

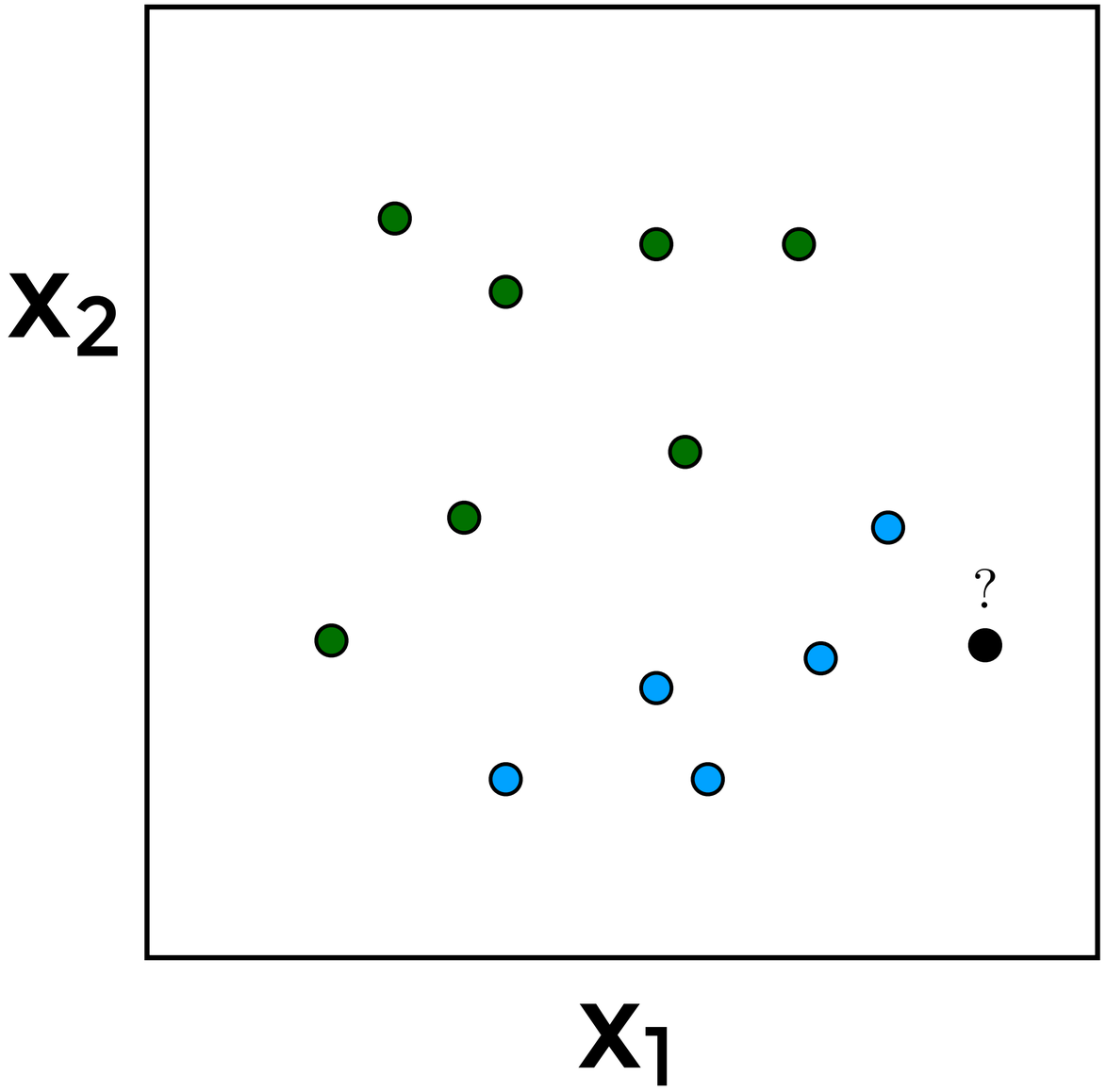
Task

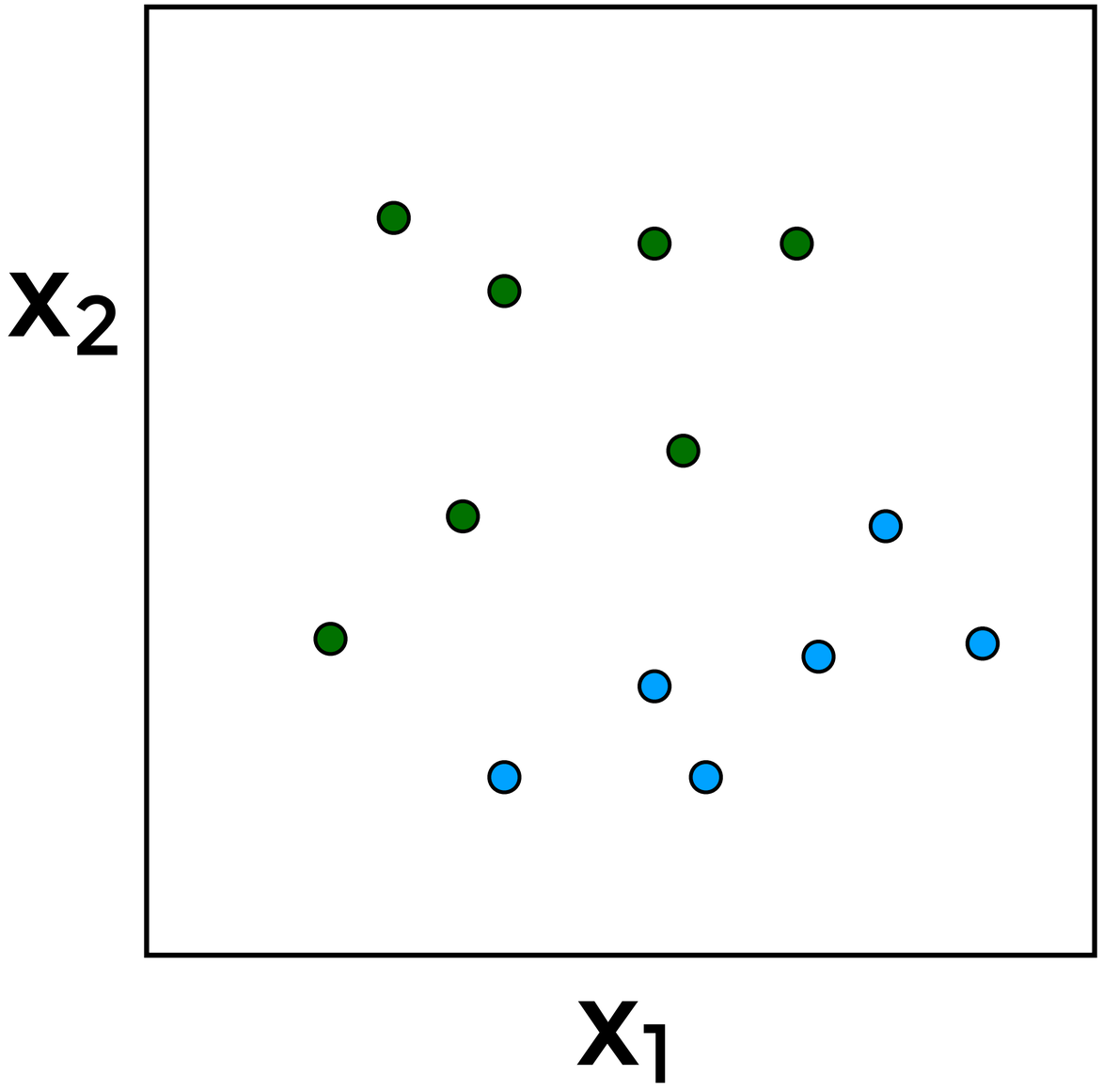
Models f $\mathbb{R}^n \rightarrow \{0, 1, 2\}$

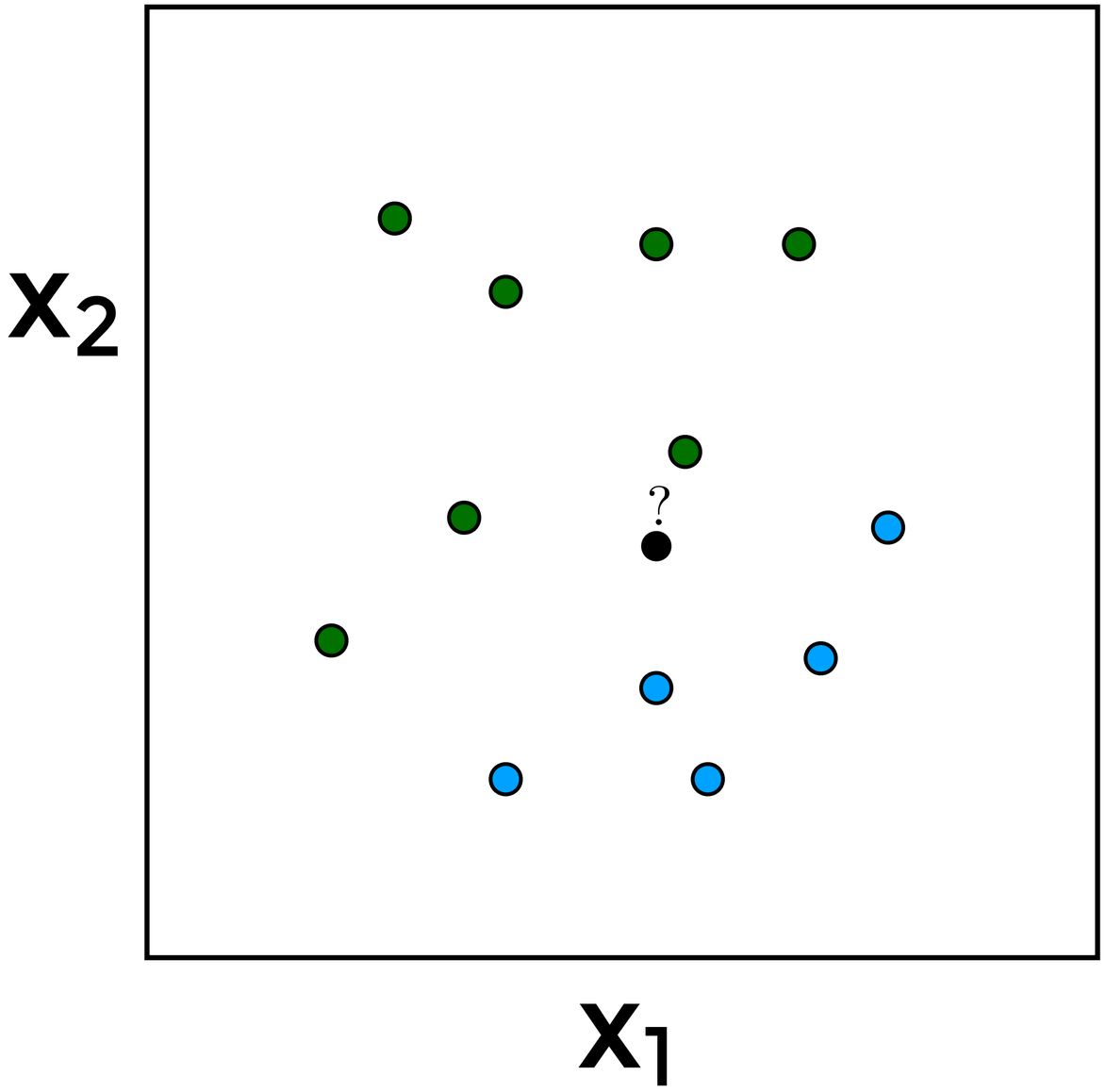
Test Data

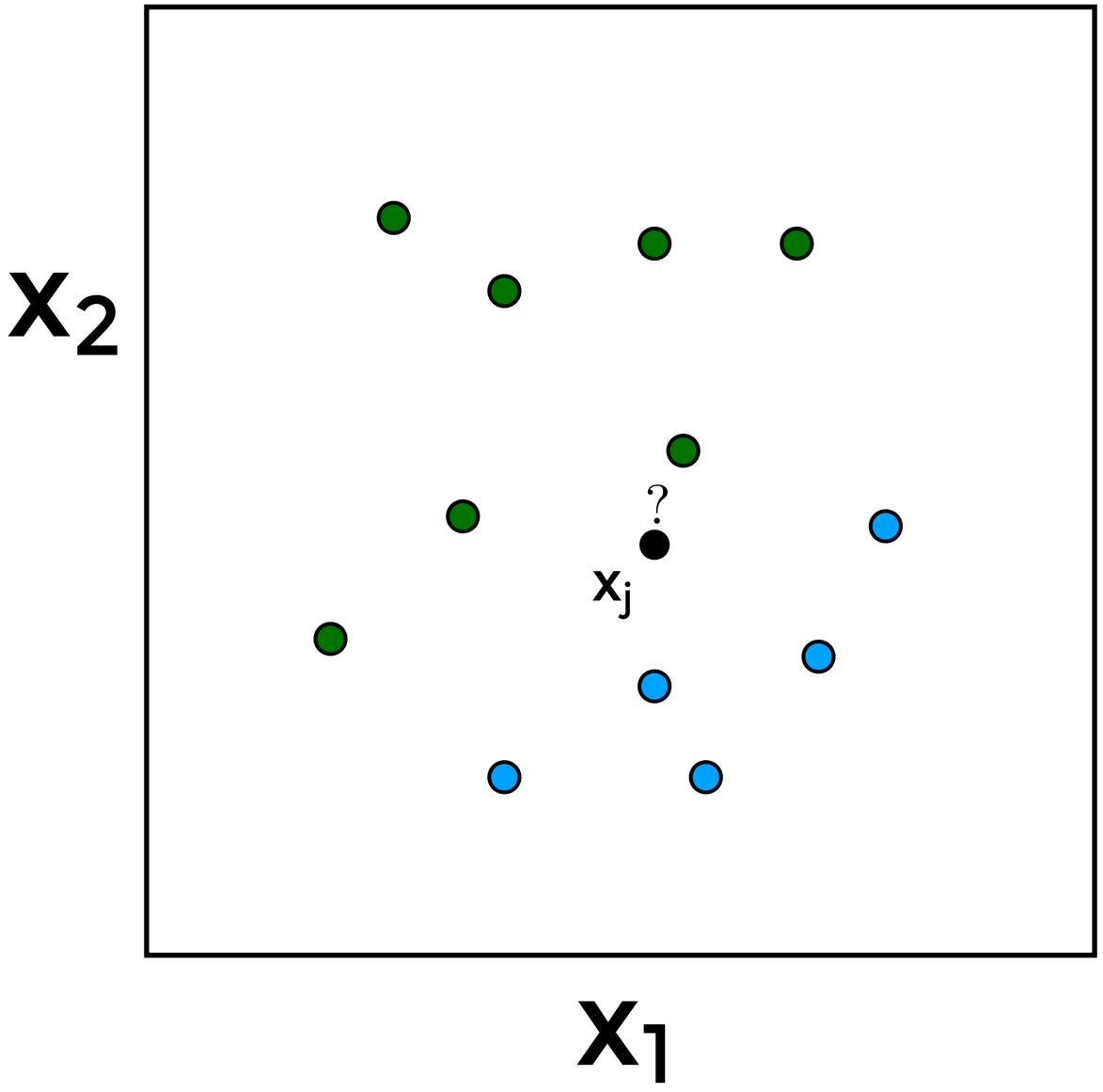
	Nb.bed.	Area	Neigh.	·	·		Sell-ability
x_0	1	0	0	0	0	y_0	?
x_1	2	50	1	.3	.8	y_1	?
x_2	1	100	1	.5	1.4	y_2	?
x_3	4	170	0	.7	.4	y_3	?
x_4	1	120	3	.9	.5	y_4	?
	X^{new}						Y^{new}


x_2
x_1


x_2
?
x_1

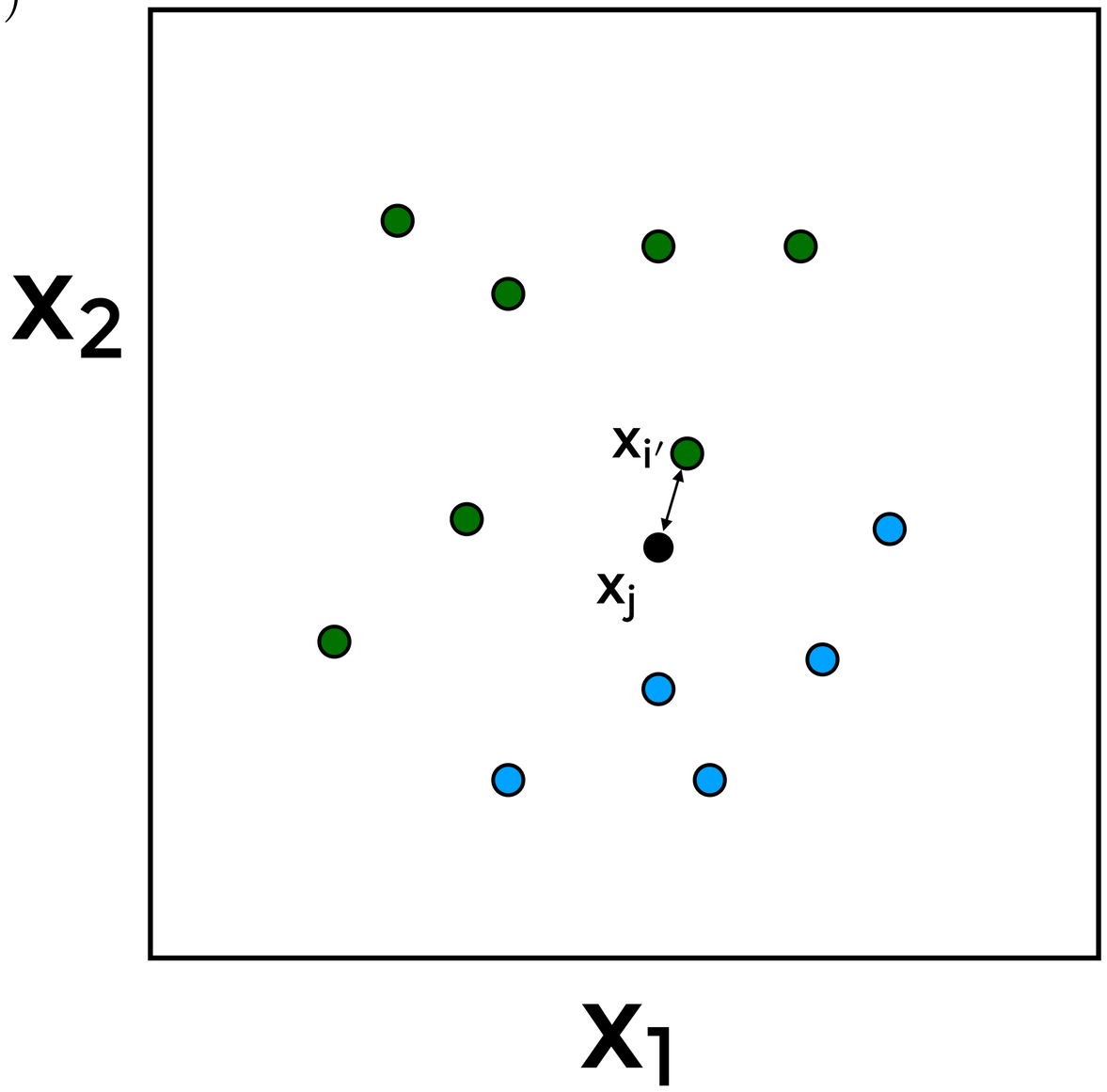

x_2
x_1


x_2
?
x_1


x_2
?
x_j
x_1

$$i' = \arg\min_i \mathbf{dist}(x_i, x_j)$$

$$y_j = y_{i'}$$



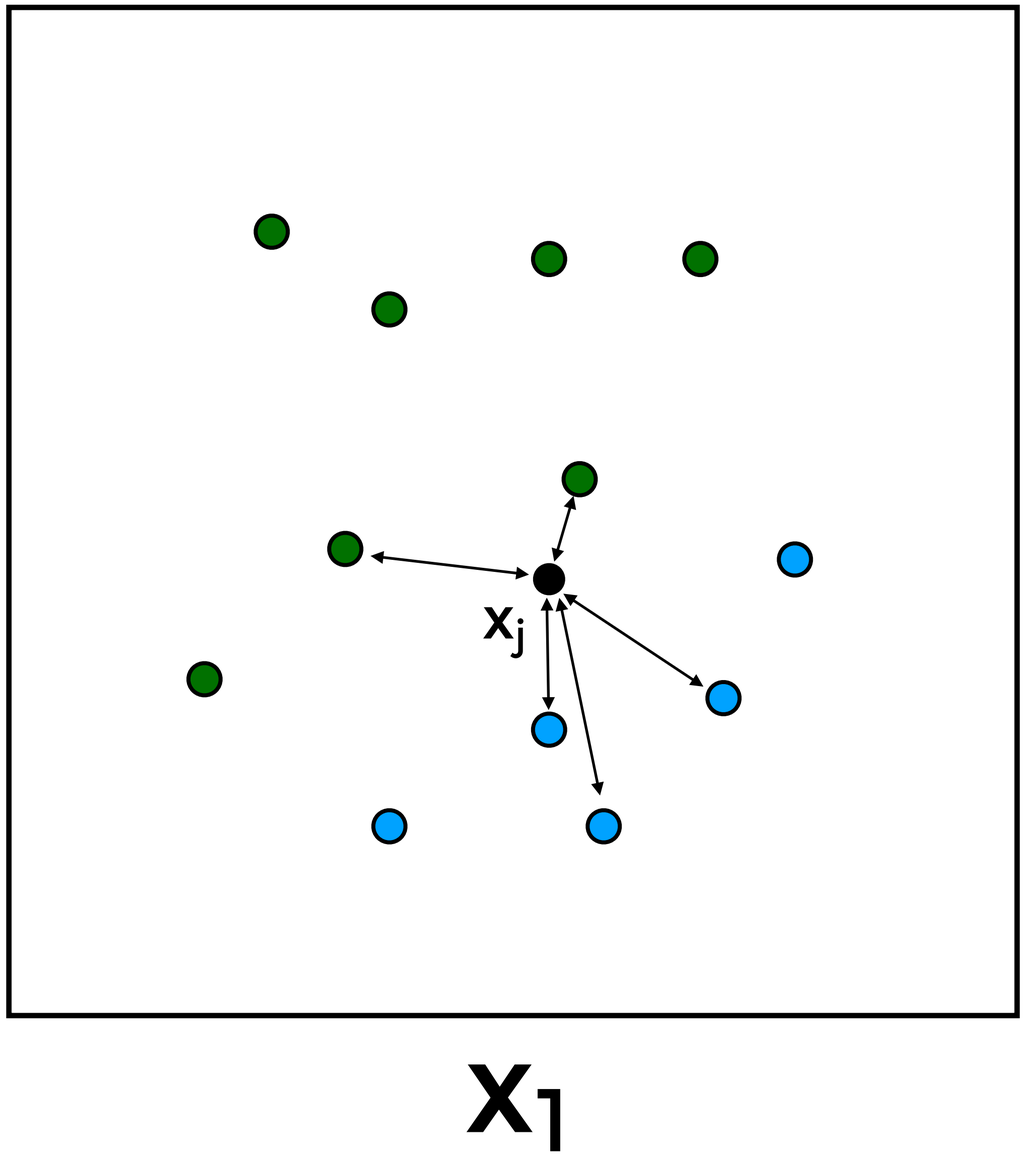
- **1-NN**

Instance classified according to its nearest neighbor

$k = 5$ (assumption)

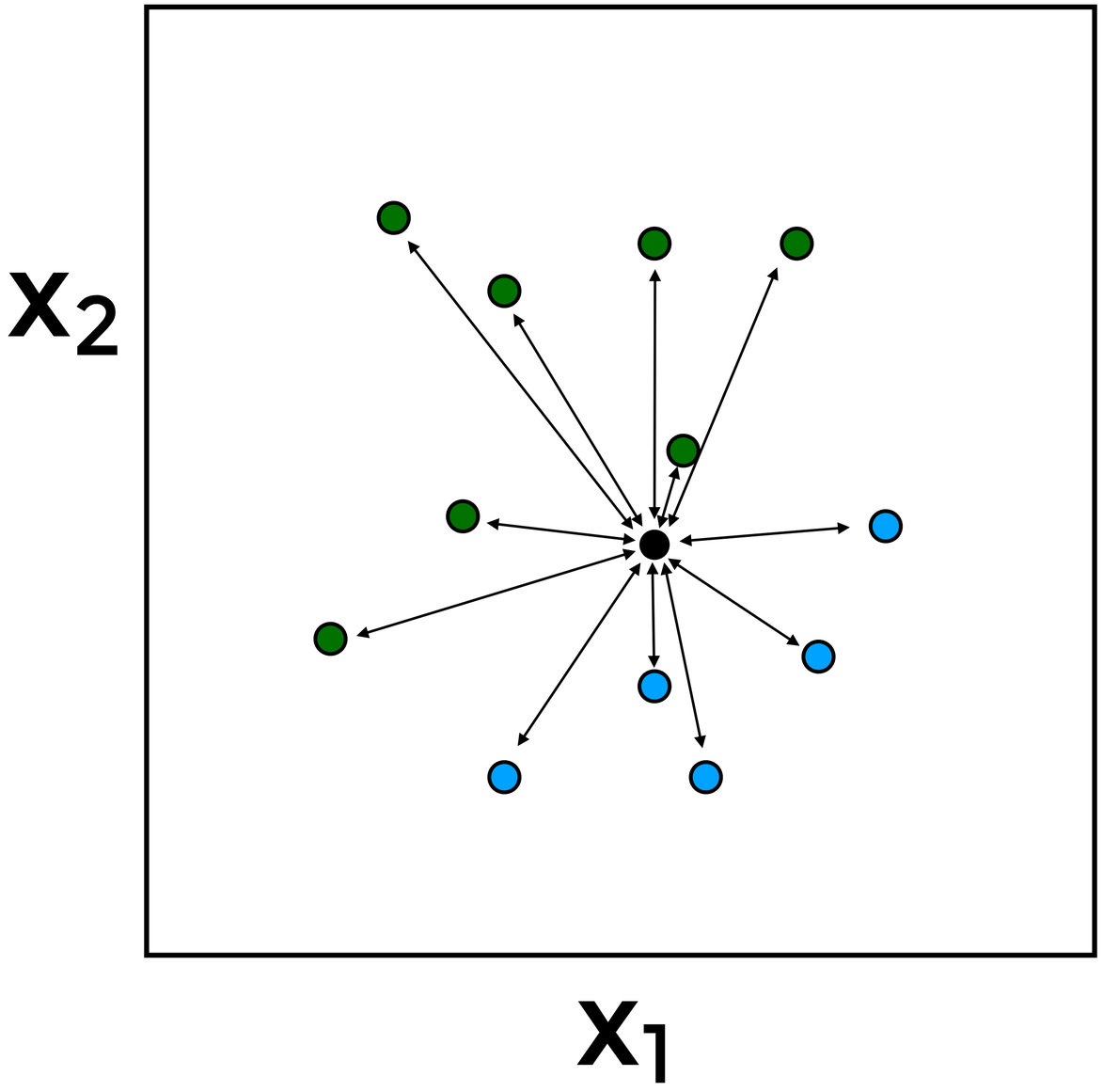
$i = \arg\text{sort}_i \, \text{dist}(x_i, x_j)$

$y_j = \text{majority}(i_{:5})$



- **K-NN**

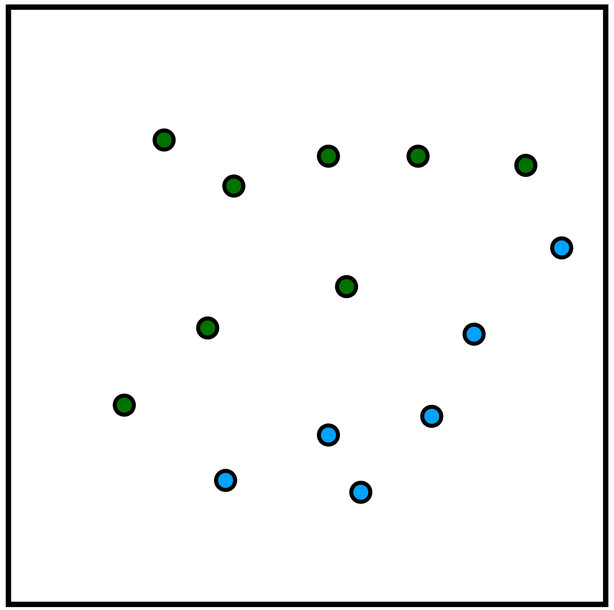
Instance classified according to the majority of its K nearest neighbors



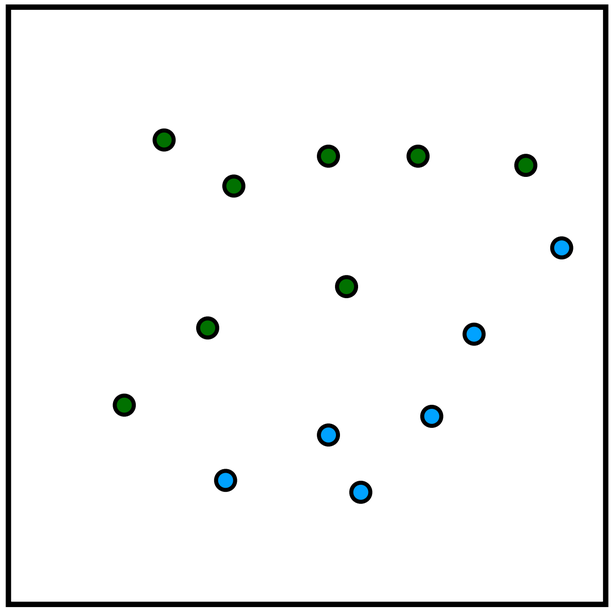
- **weighted-NN**

Instance classified according to all neighbors. The contribution of each neighbor is weighted by its distance.

Linear Classification

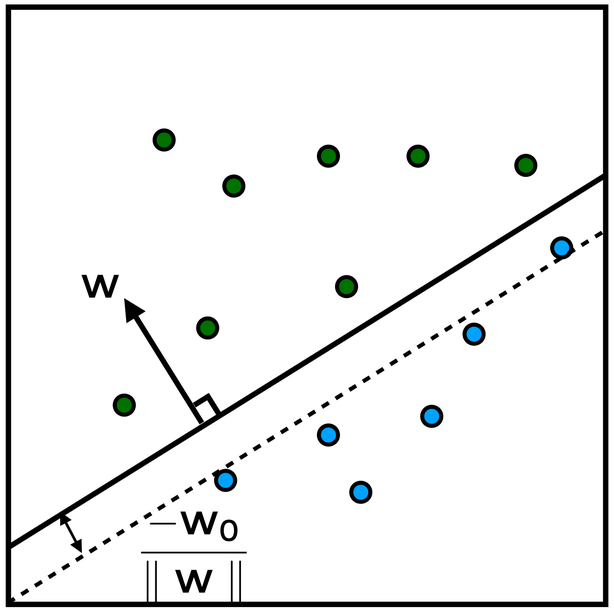


Linear Classification



$$\mathbf{y}(\mathbf{x}) = \mathbf{w}^\top \mathbf{x} + \mathbf{w}_0$$

Linear Classification

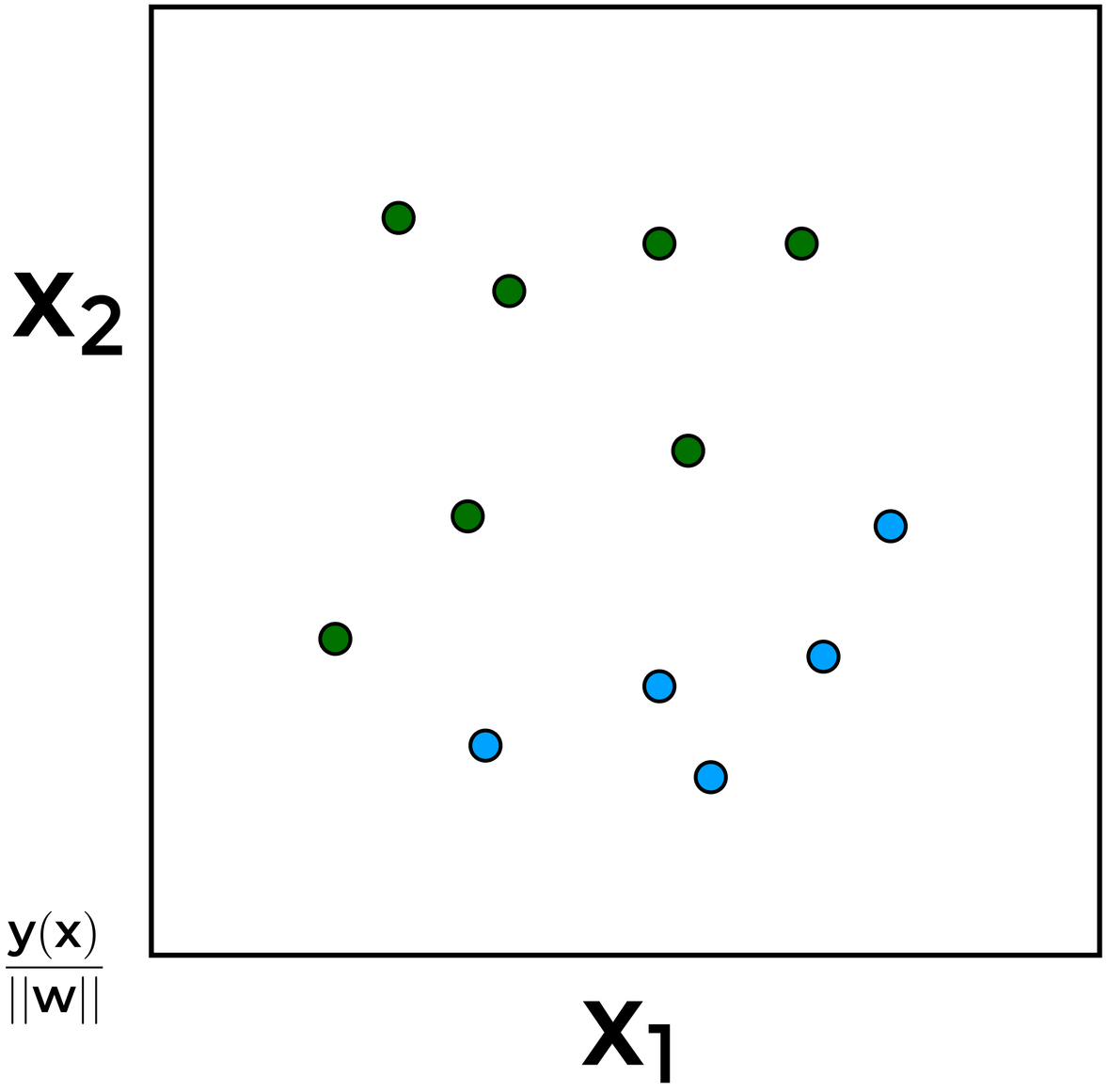


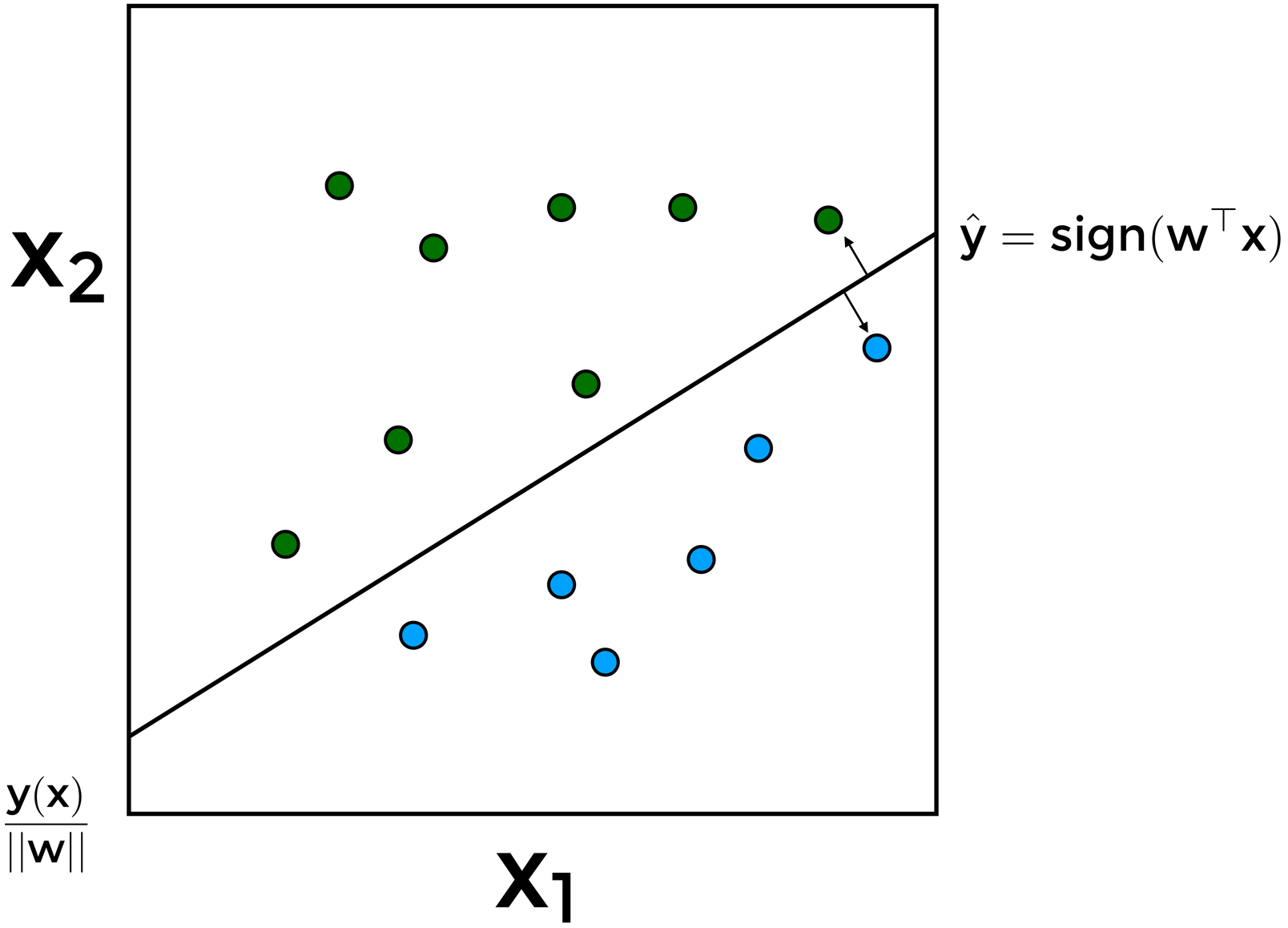
$$y(x) = w^\top x + w_0$$

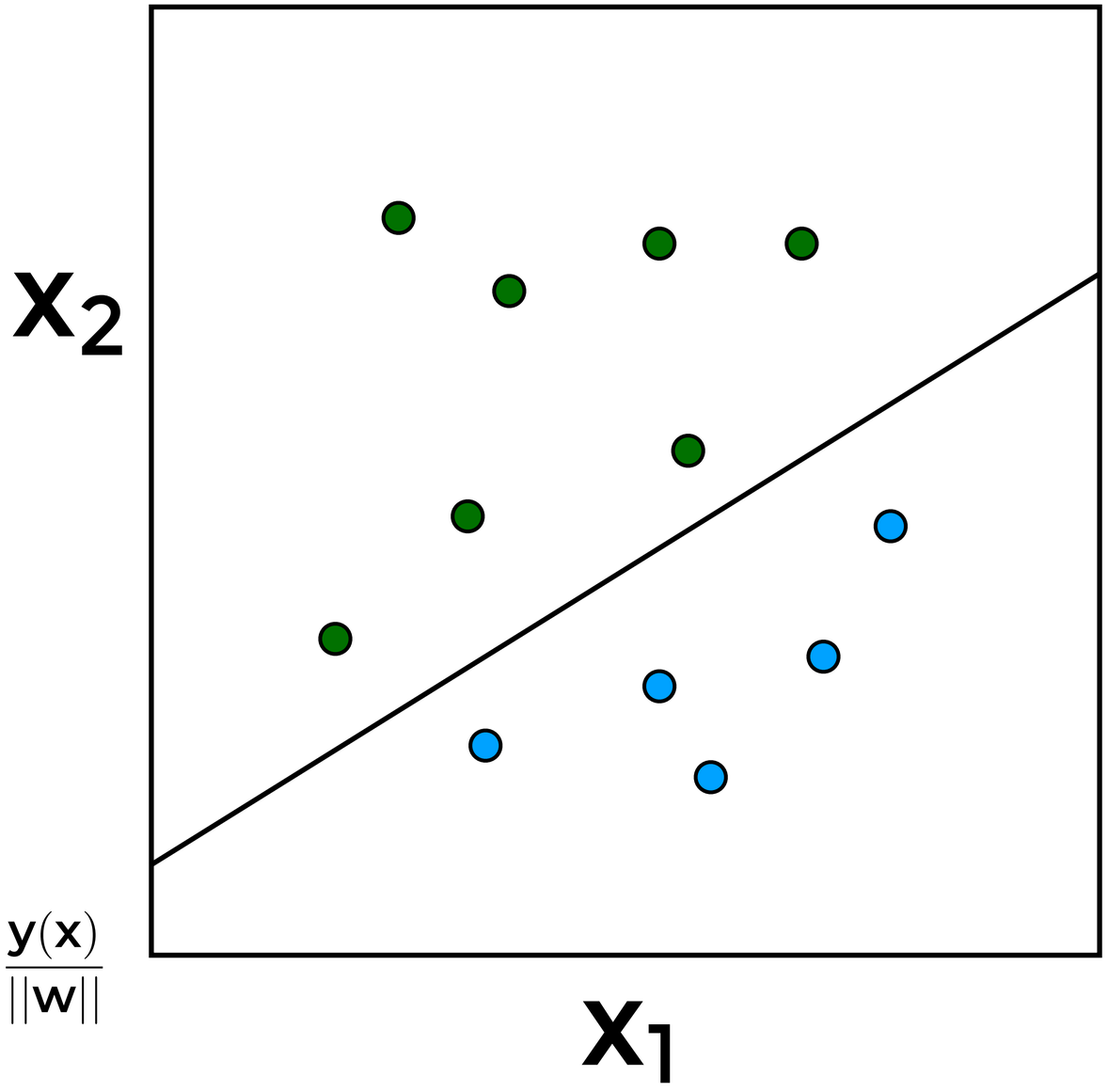
Decision

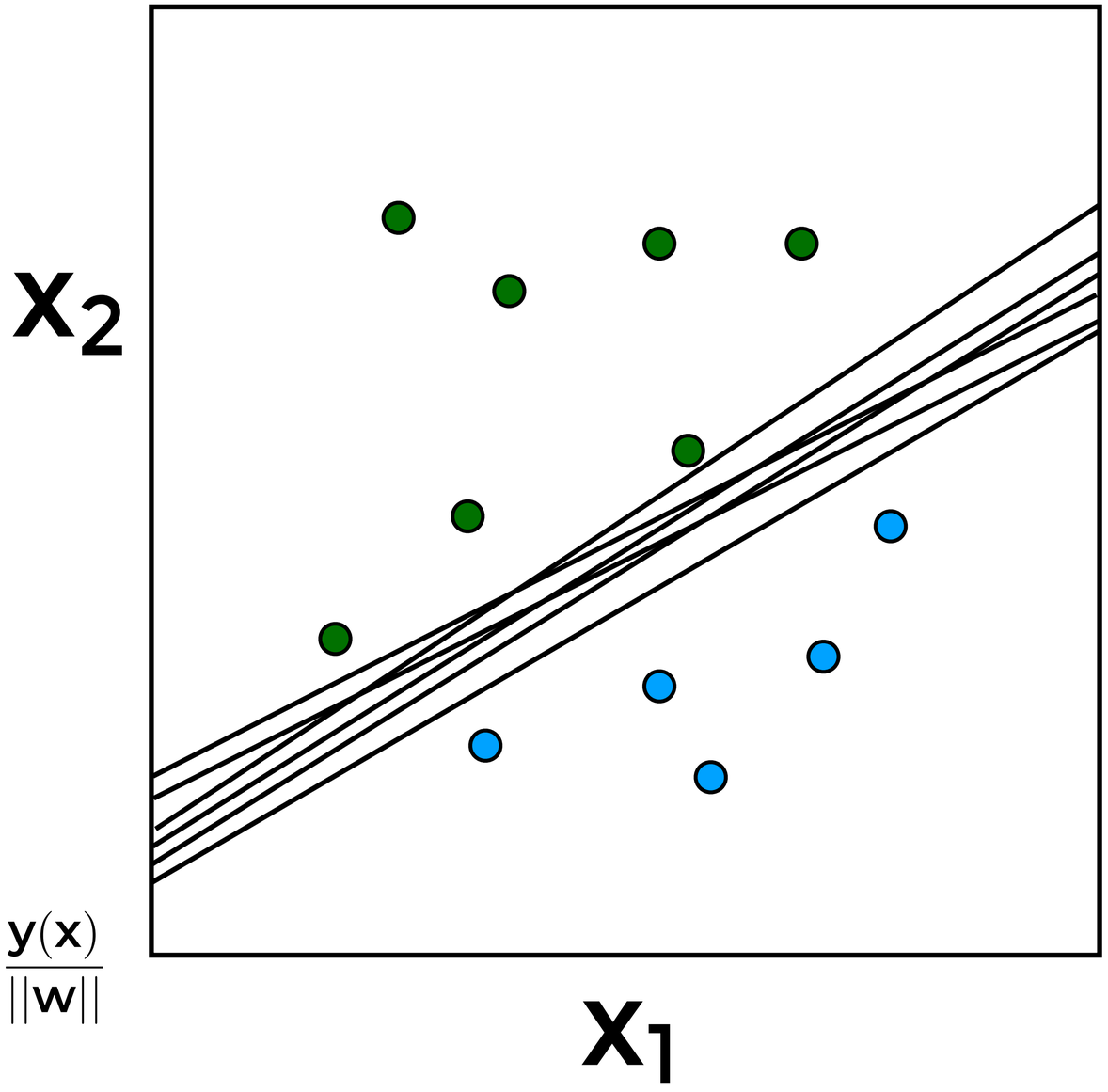
$$(w^\top x + w_0) > 0 \implies$$

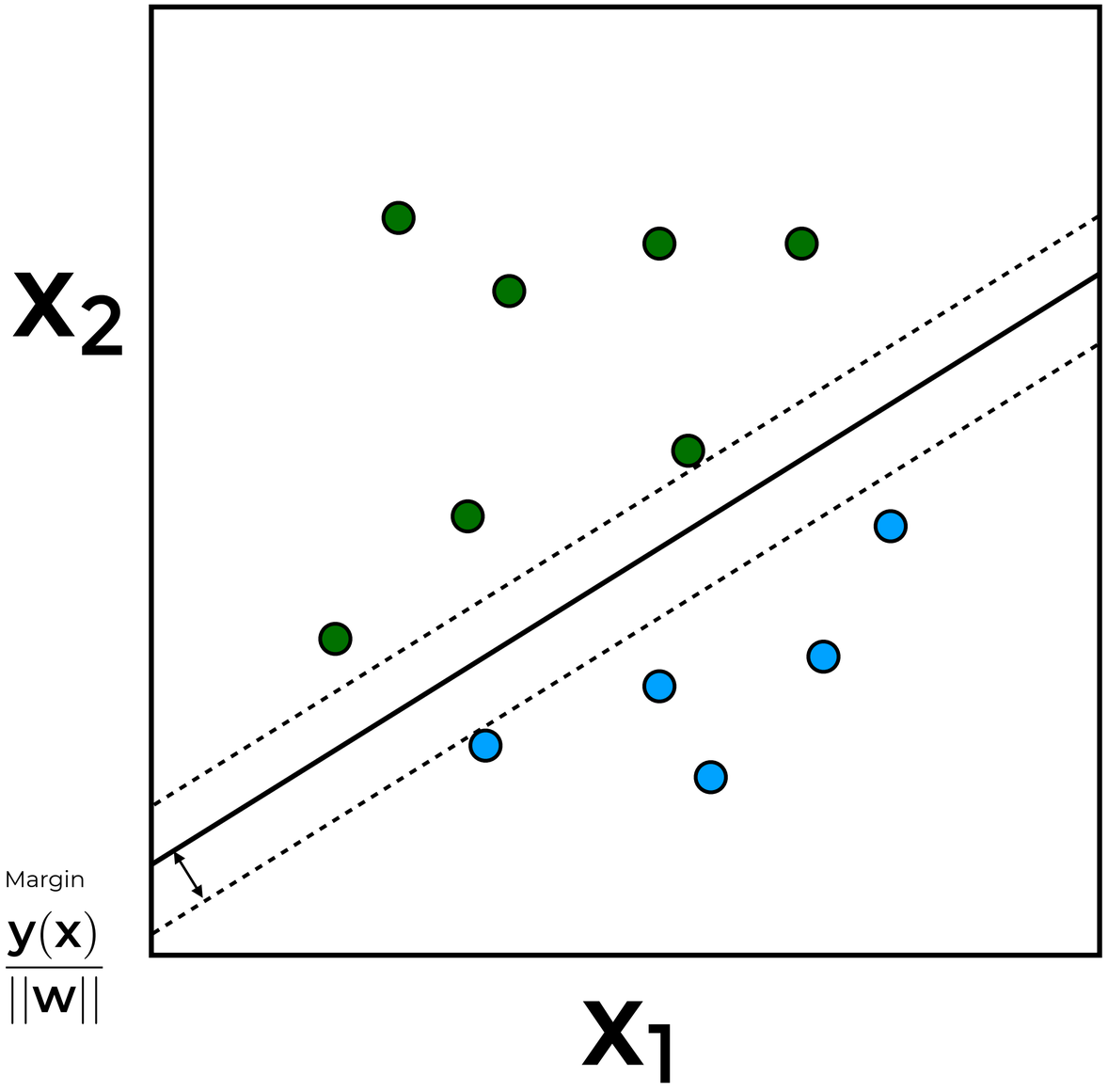
$$(w^\top x + w_0) < 0 \implies$$


x_2
$\frac{y(x)}{||w||}$
x_1

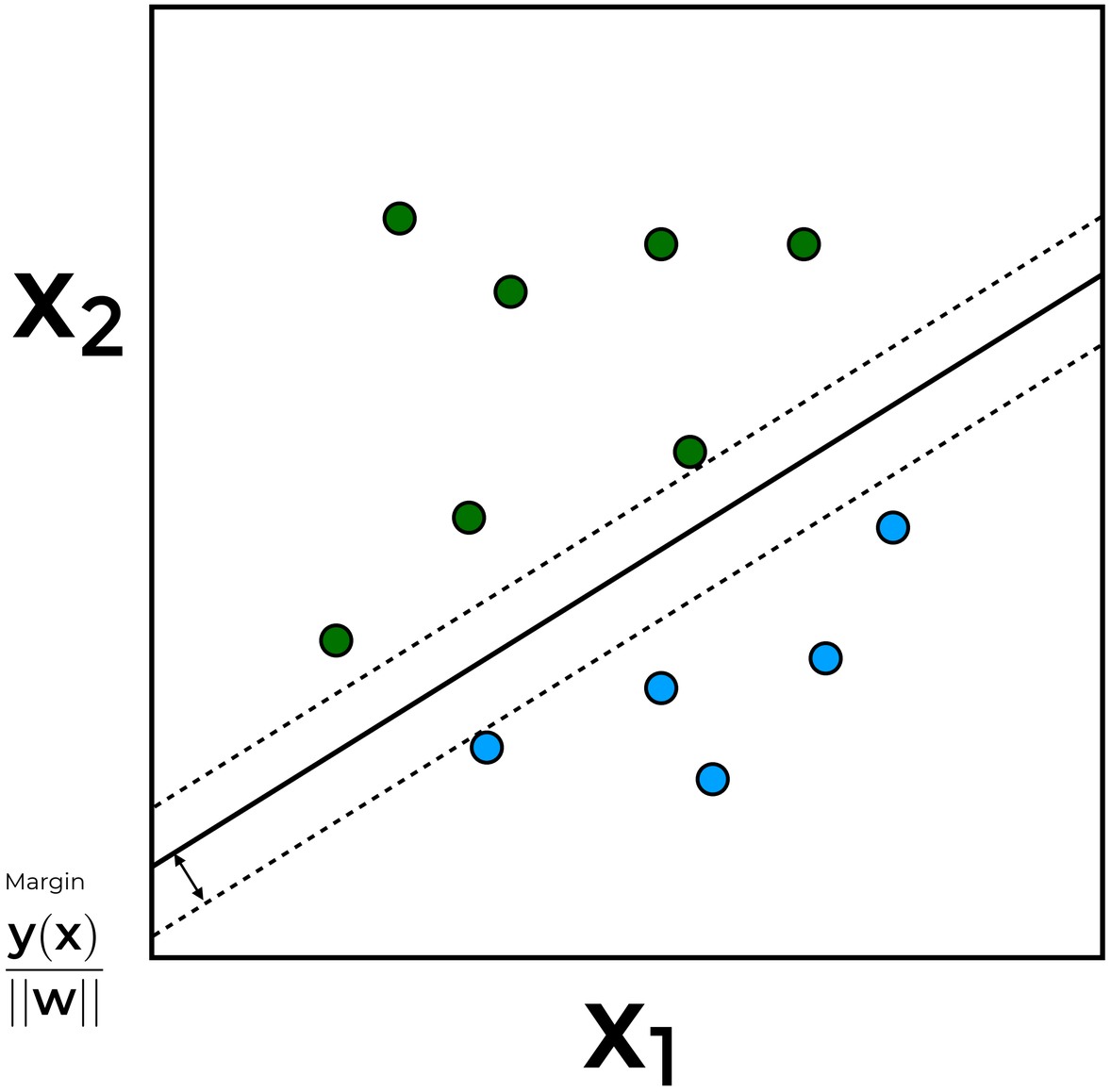

$\hat{y} = \text{sign}(w^\top x)$
x_2
x_1
$\frac{y(x)}{||w||}$


x_2
$\frac{y(x)}{||w||}$
x_1


x_2
$\frac{y(x)}{||w||}$
x_1

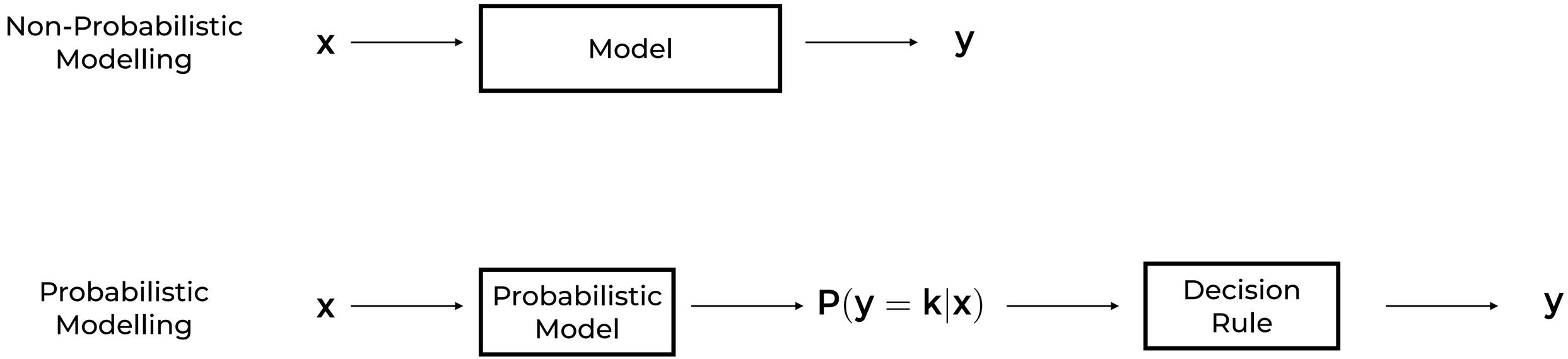

x_2
x_1
Margin
$\frac{y(x)}{||w||}$

The objective is to find the separating boundary that maximizes the margin



Probabilistic Models for Classification

Probabilistic Models separate Decision and Inference



Probabilistic Modelling

Often inteertwinned

1. Posit a model: P(X, Y)
   - How the data is generated
2. Parametrize the distributions: P(X, Y | Parameters)
3. Set the objective (e.g., MLE)
4. Learn the parameters of the model:
   - E.g., Naive Bayes: learn the parameters of the class conditional P(X | Y) and of the prior P(Y)
5. Use the model (e.g., for predictions)