Week 9

Machine Learning

Leo Breiman's Two Cultures

the logic of data analysis

X → Nature → **y**

input parameters

response variables

L. Breiman, Statistical Science (2001)

Leo Breiman’s Two Cultures

Data Modeling Culture

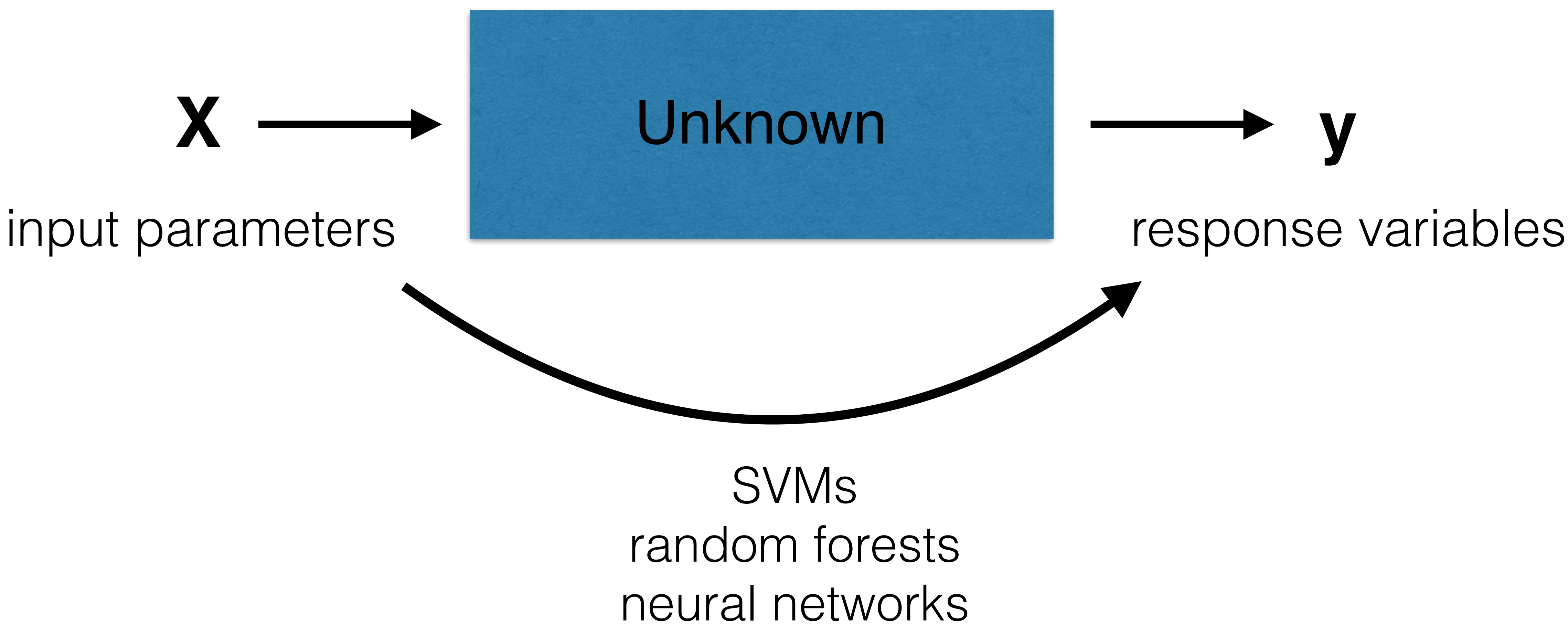
e.g. linear regression

Focus on stochastic model to explain
how f(x)-> y

98% of Statistics

Leo Breiman’s Two Cultures

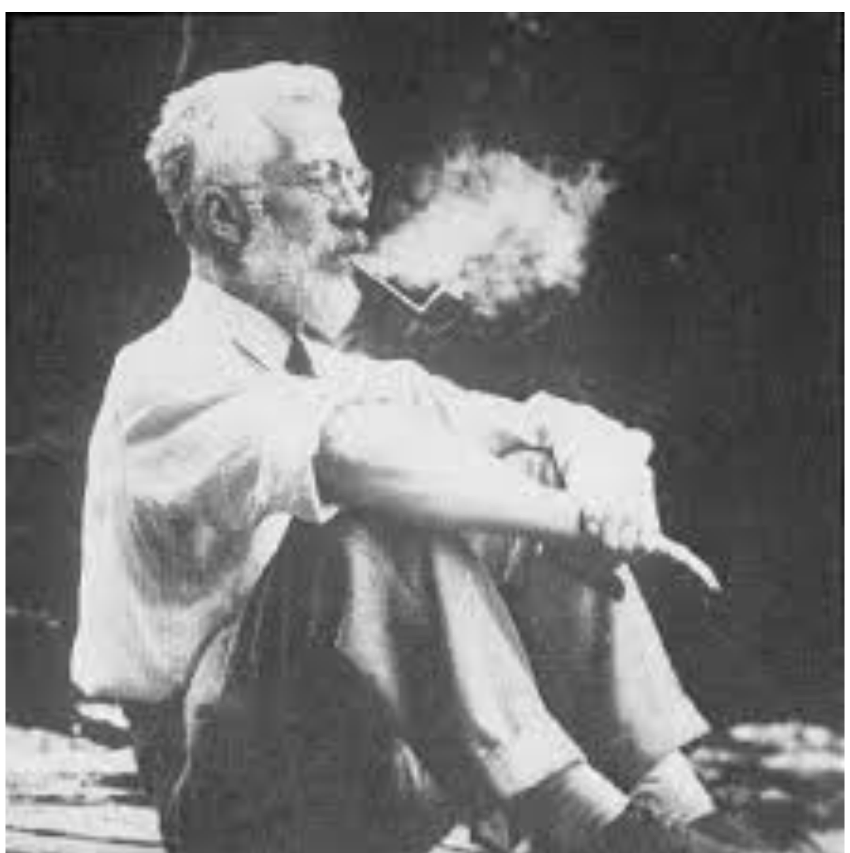
Algorithmic Modeling Culture
(machine learning)

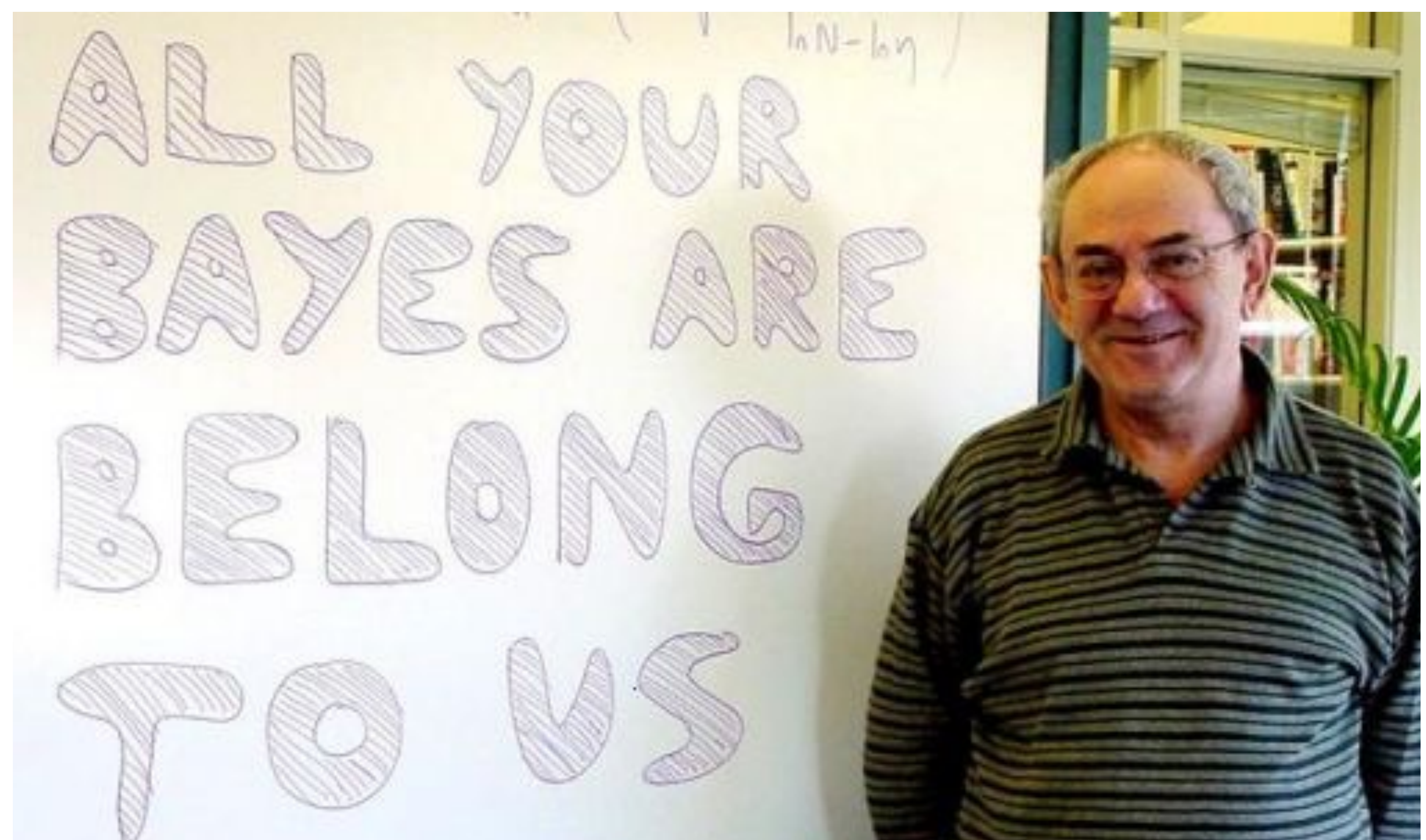


Ignore probabilistic generative model f(x)-> y

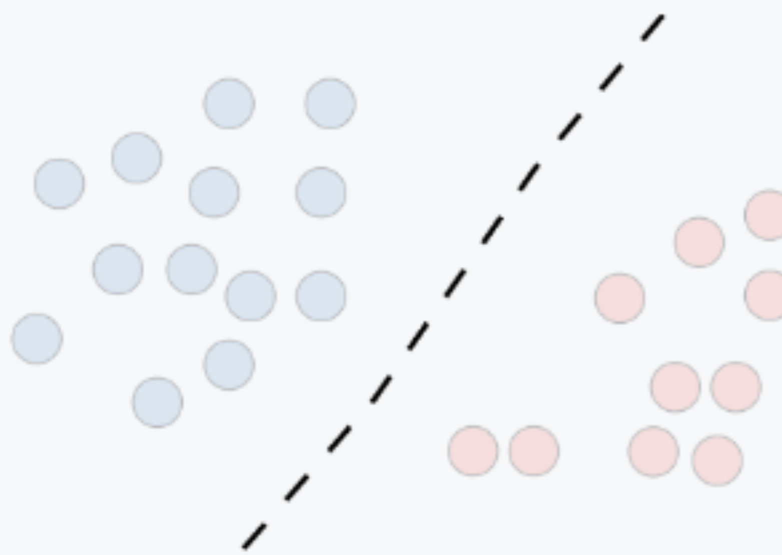
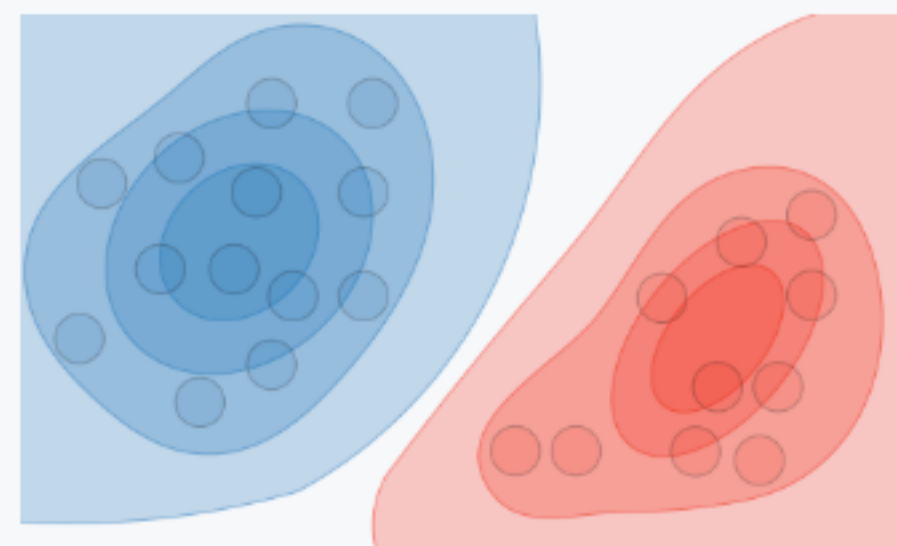
Machine Learning!

These guys don’t have generative model

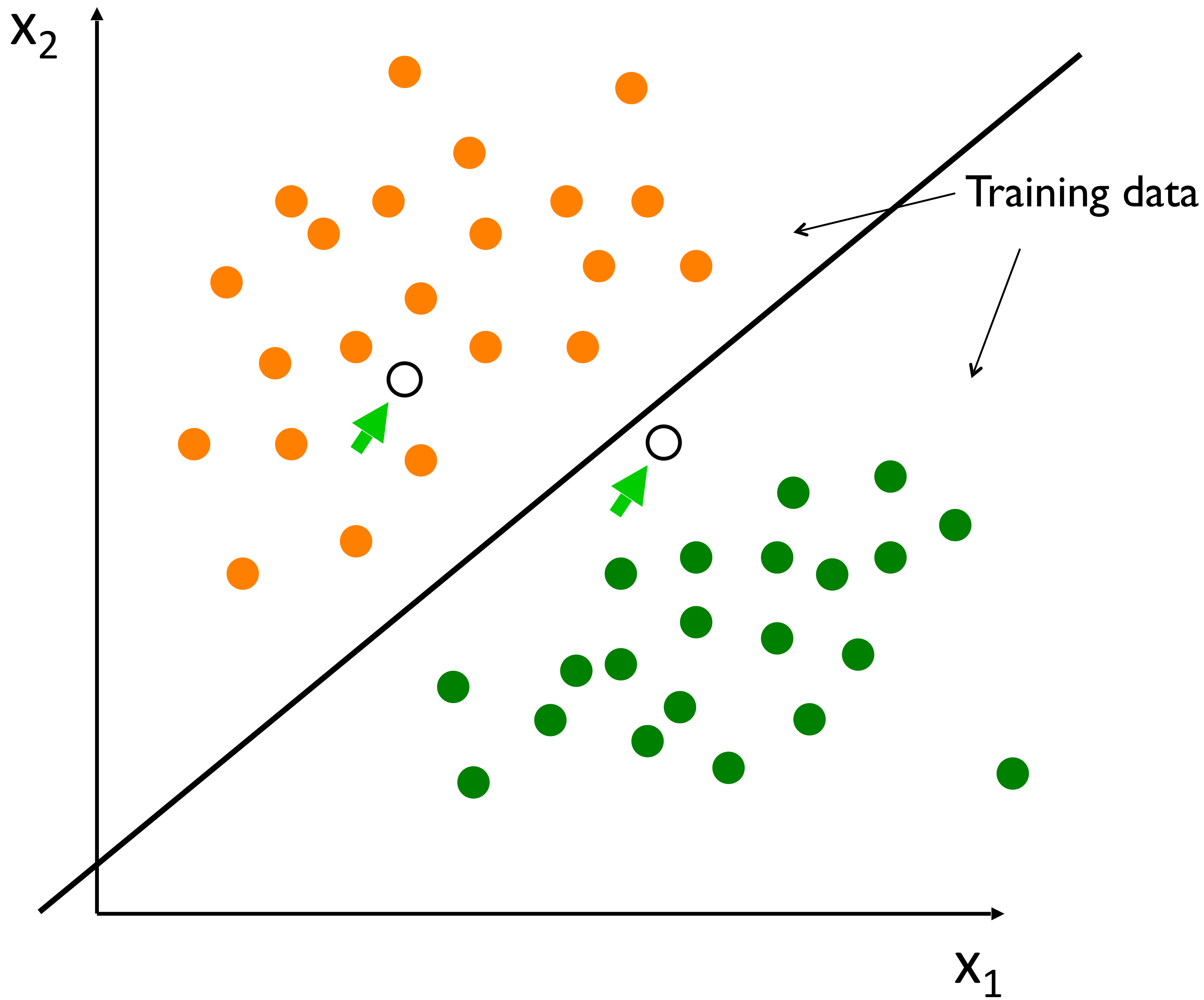




Discriminitive vs Generative Models

	Discriminative model	Generative model
Goal	Directly estimate $P(y\|x)$	Estimate $P(x\|y)$ to then deduce $P(y\|x)$
What's learned	Decision boundary	Probability distributions of the data
Illustration		
Examples	Regressions, SVMs	GDA, Naive Bayes

Supervised Machine Learning



We are using a Support Vector Machine (SVM)

Supervised Machine Learning

Given a set of *N* training (i.e. known, labelled) examples:

$$\{(x_1, y_1), ..., (x_N, y_N)\}$$

feature vector $\mathbb{R}^M$ class label $y \in \{-1, 1\}$

we define a learning function:

$$g : X \to Y \quad \text{e.g.} \quad g(x) = P(y|x)$$

and a loss function:

$$L : g(x) \times Y \to \mathbb{R}^{\geq 0} \quad \text{e.g.} \quad L(g(x), y) = \mathbb{1}(g(x) \neq y)$$

then simply minimize a chosen risk function:

$$R(g) = \frac{1}{N} \sum_i L(y_i, g(x_i))$$

Supervised Machine Learning

Support Vector Machines

general learning function:

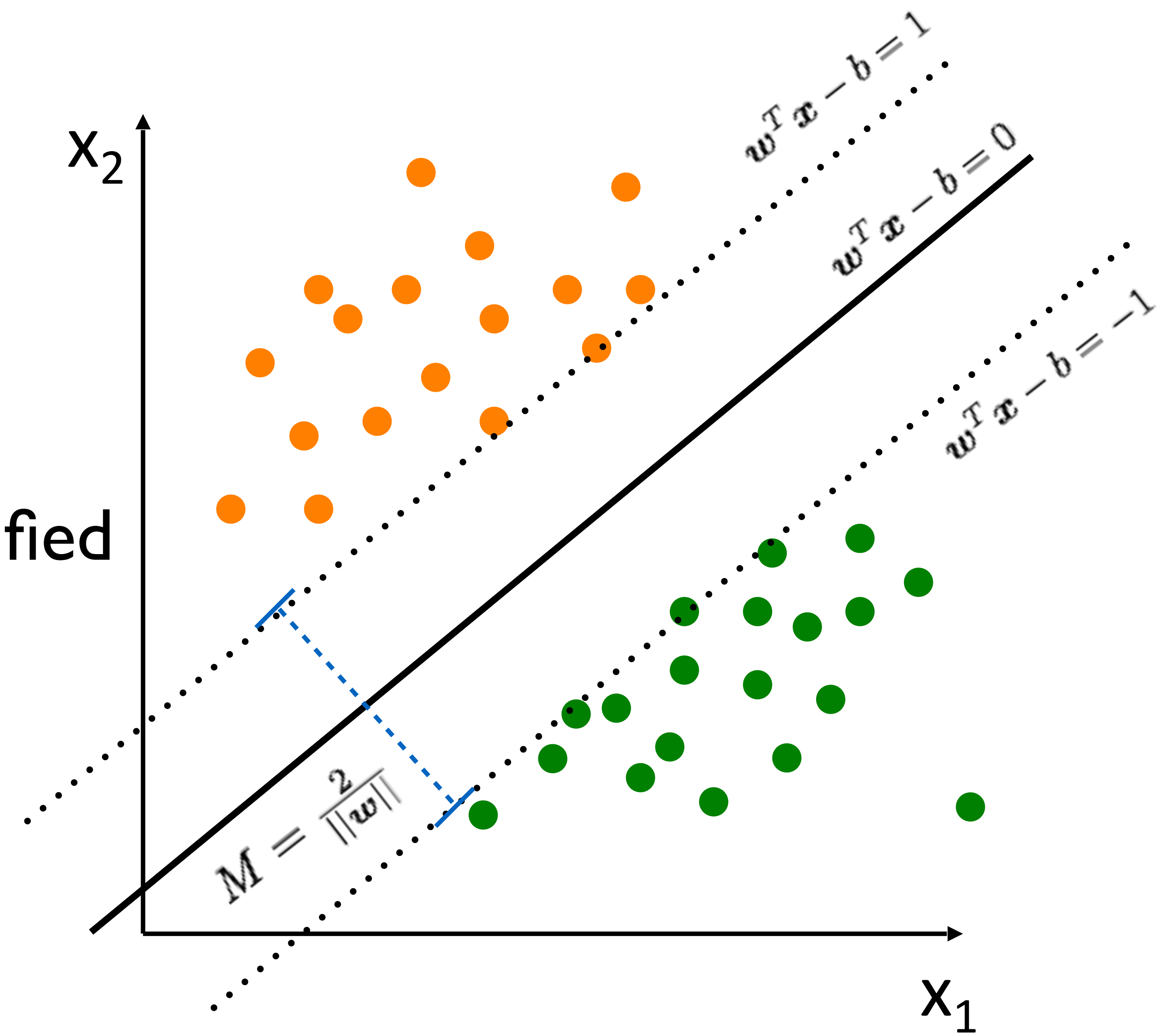
$$g(\boldsymbol{x}) = sign(\boldsymbol{w}^T\boldsymbol{x} - b)$$

simplest form = "Hard Margin"

i.e. all training points correctly classified

minimize $||\boldsymbol{w}||$ subject to,

$$y_i(\boldsymbol{w}^T\boldsymbol{x_i} - b) \geq 1$$

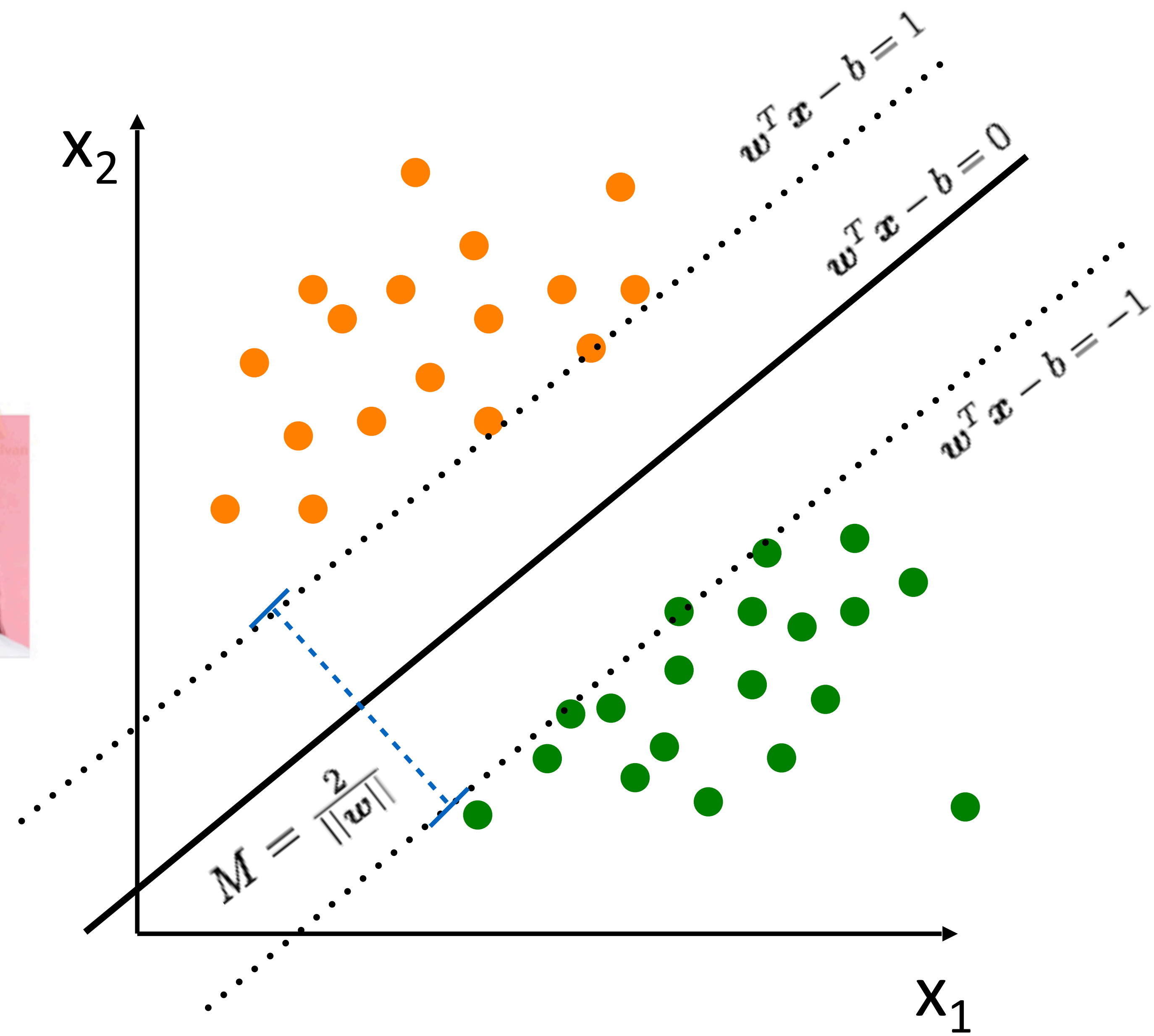


LOTS of variations on this e.g. soft margins, kernel trick for non-linear

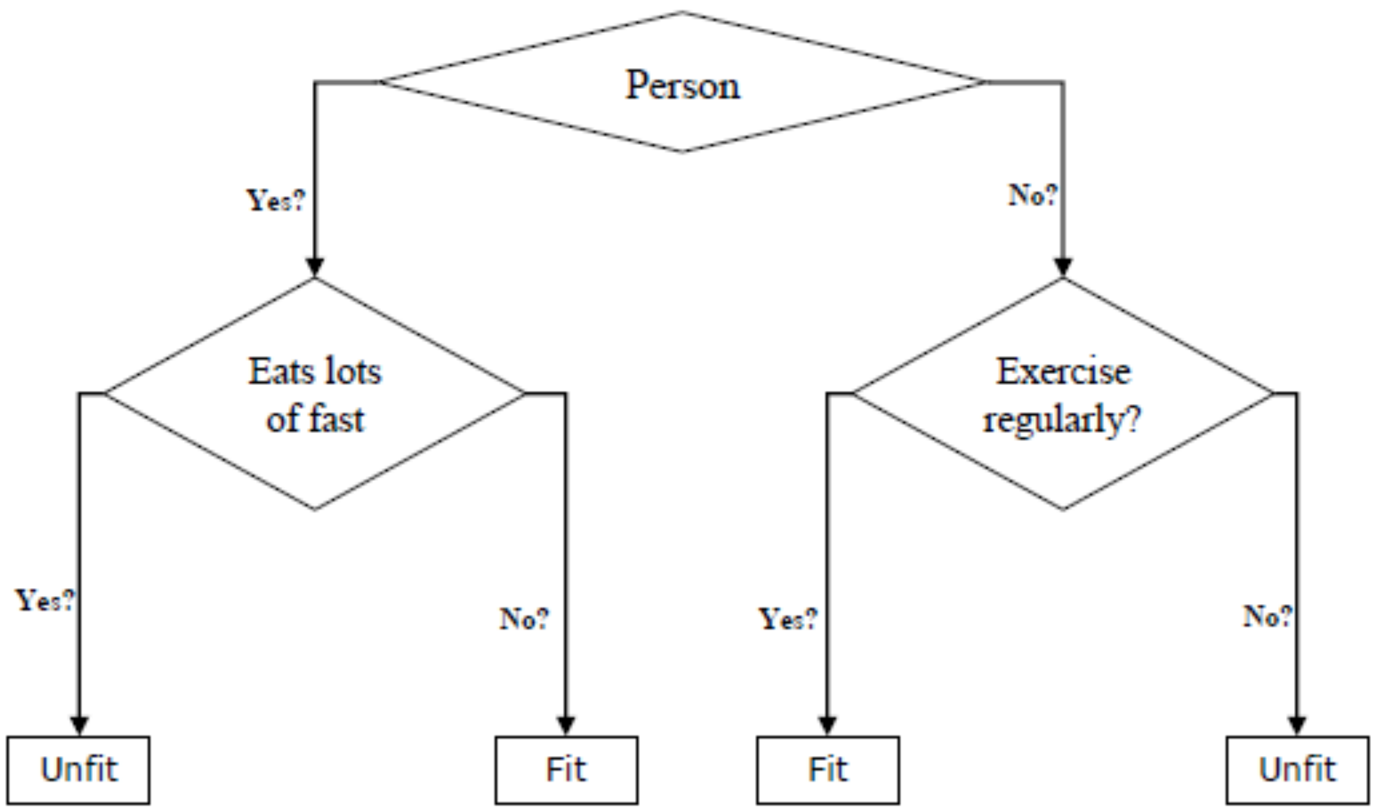
Supervised Machine Learning

Support Vector Machines

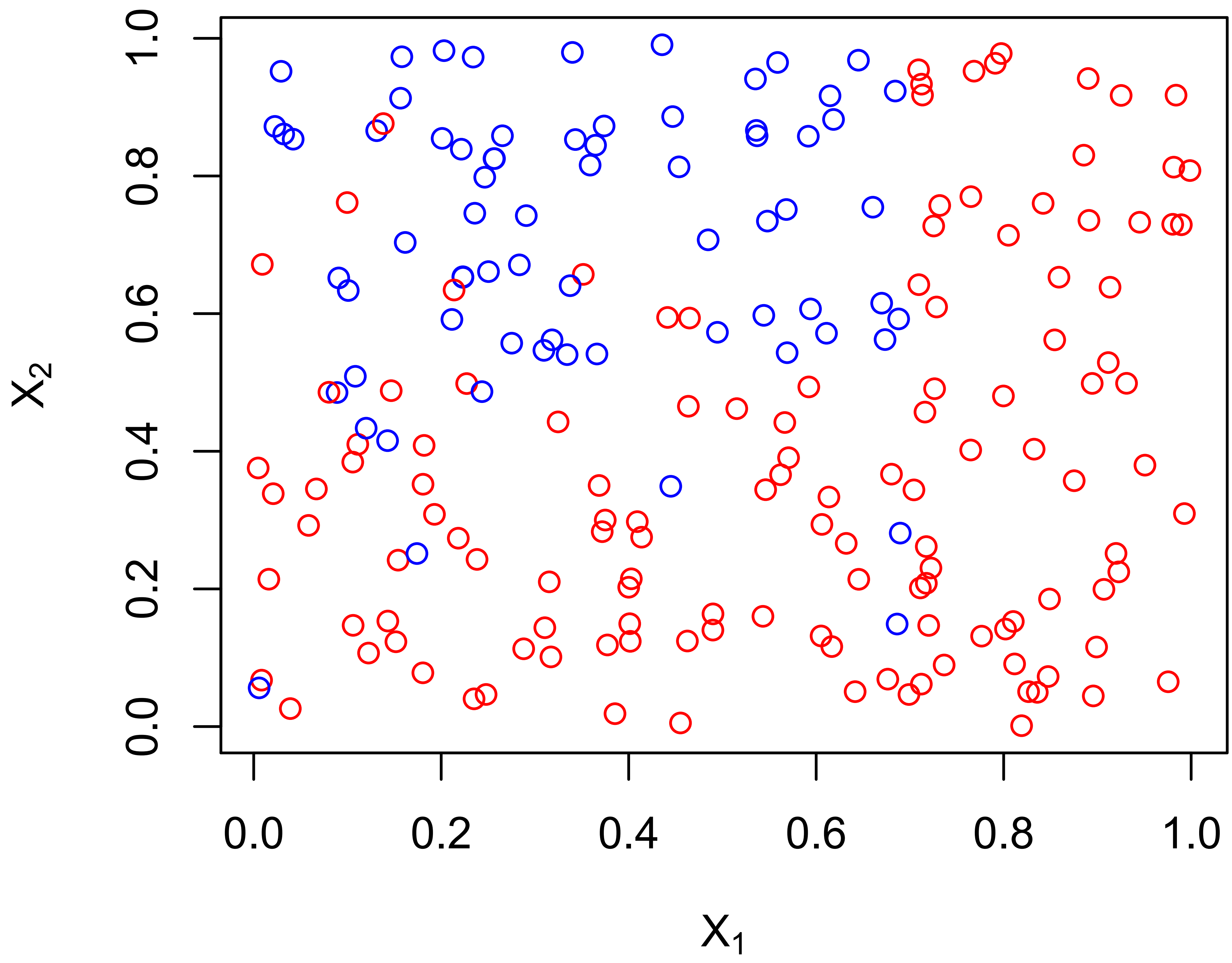
Image recognition via SVM



Decision Trees and Random Forests

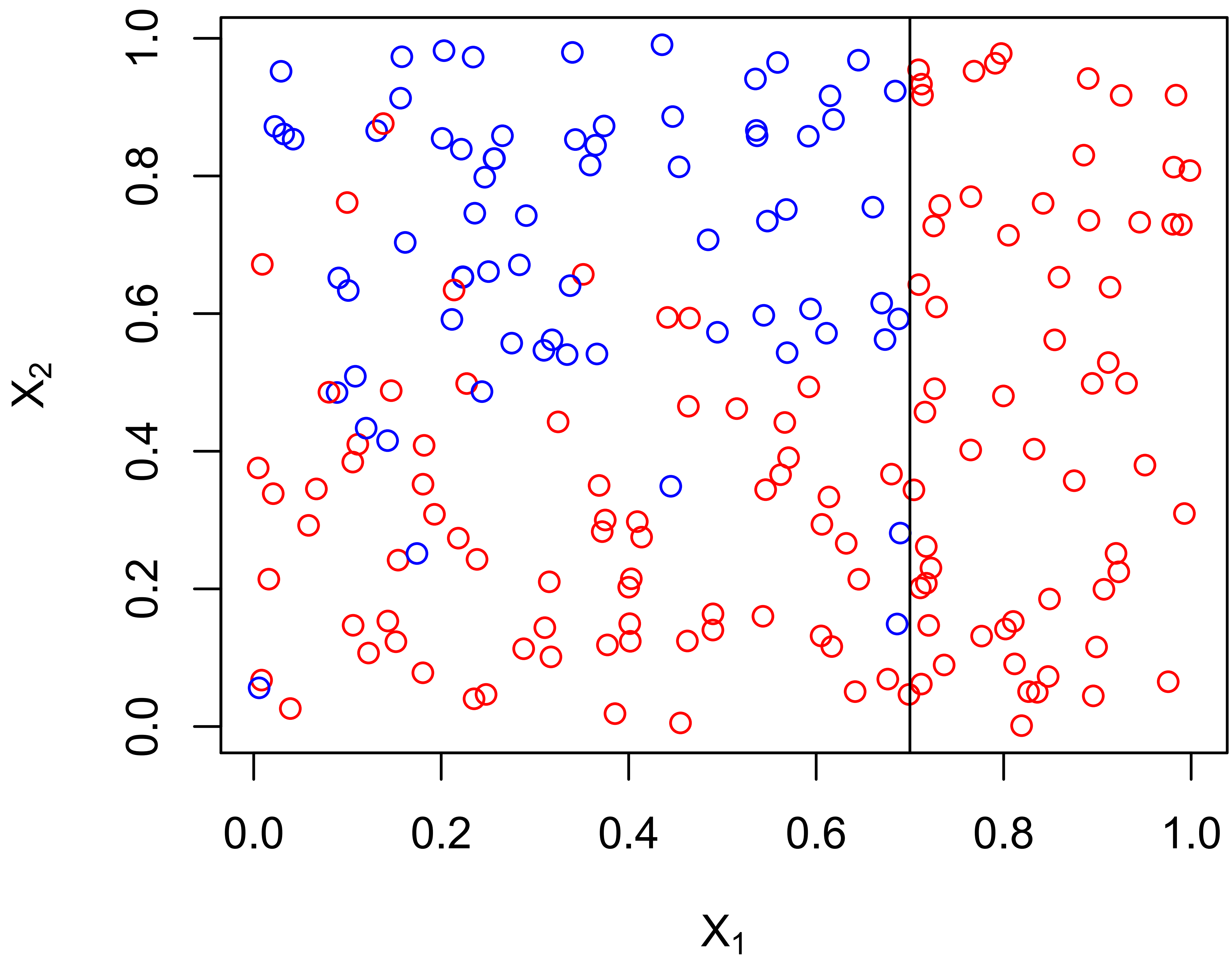


From decision trees to extra trees



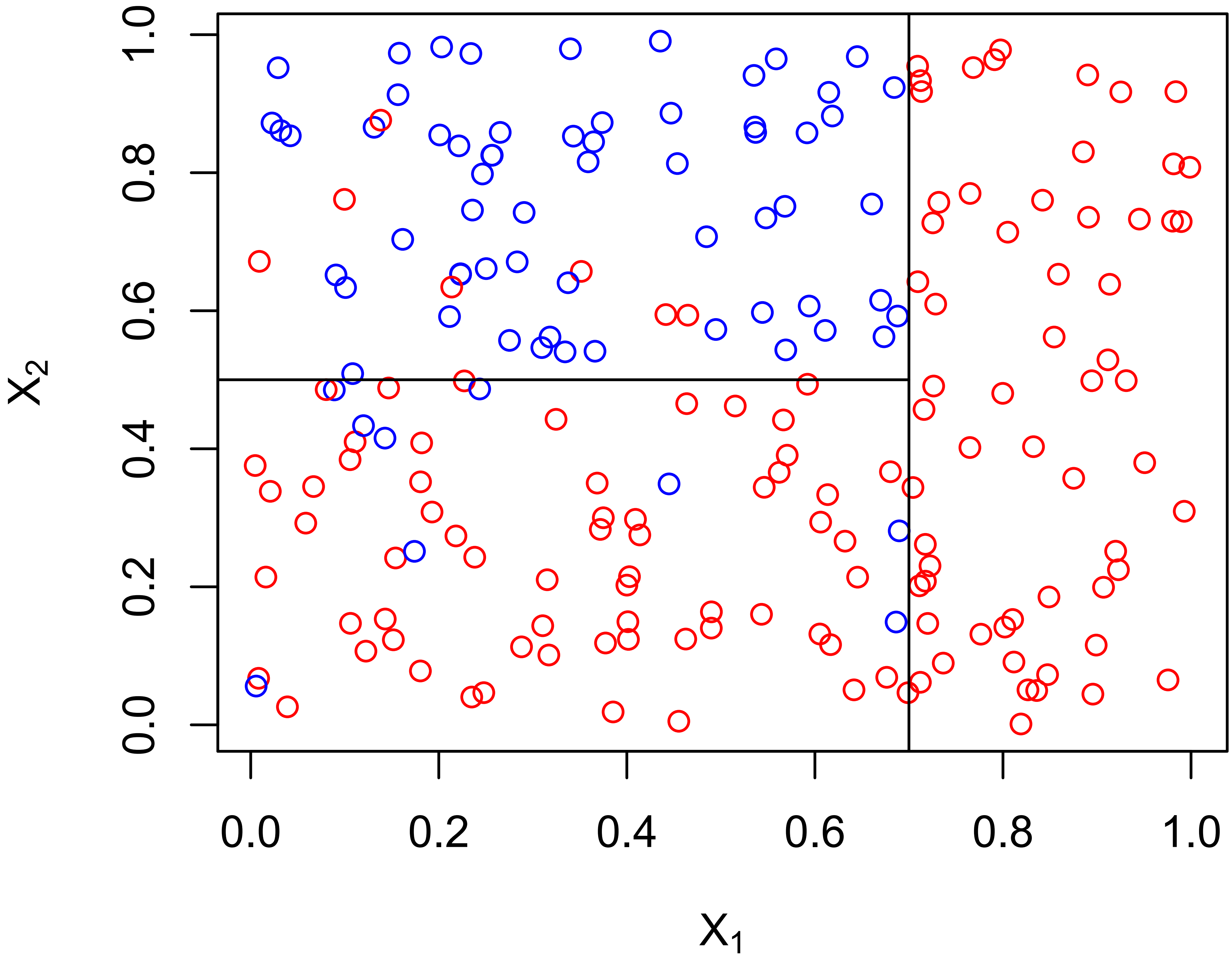
Schrider and Kern (2016)

From decision trees to extra trees



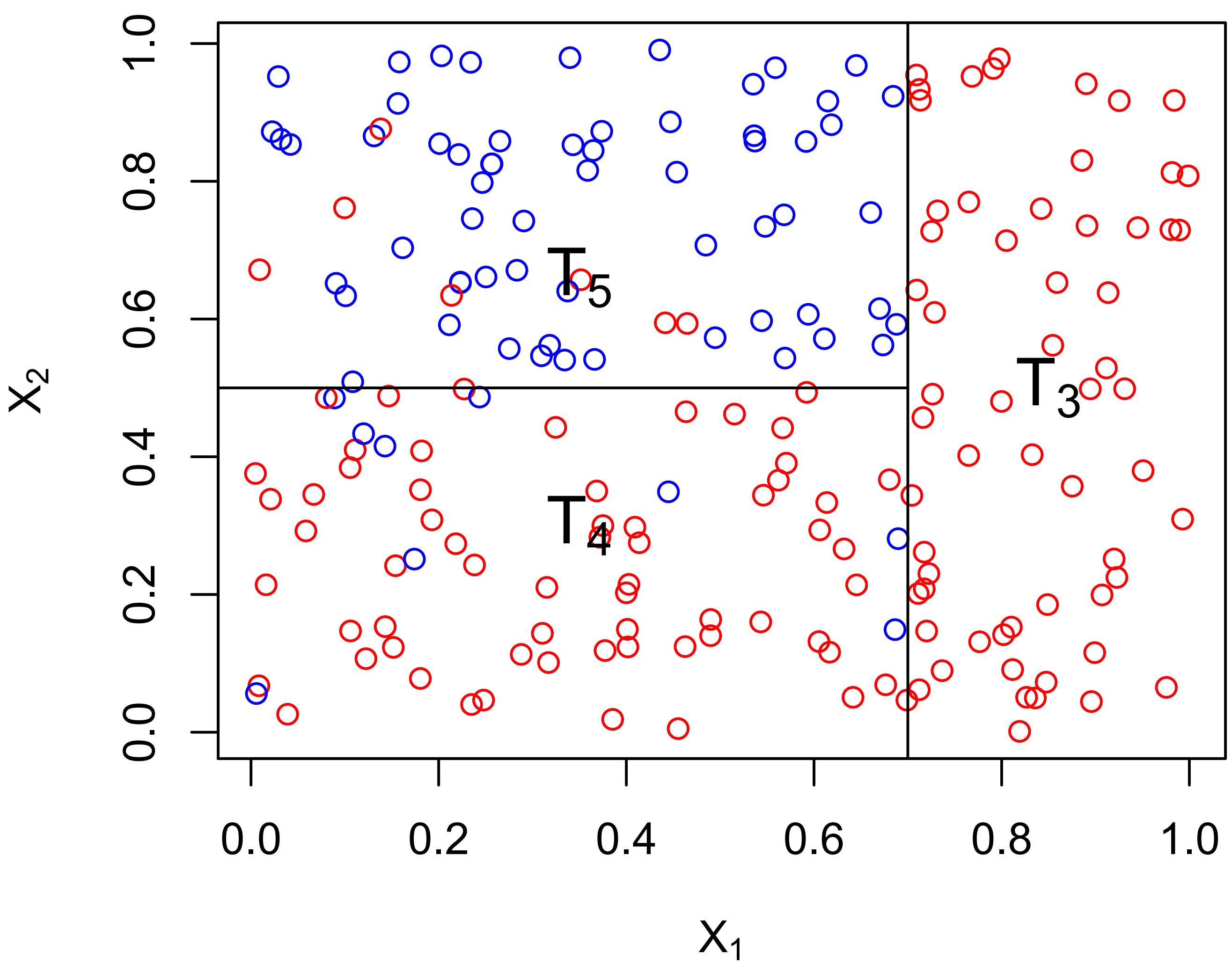
Schrider and Kern (2016)

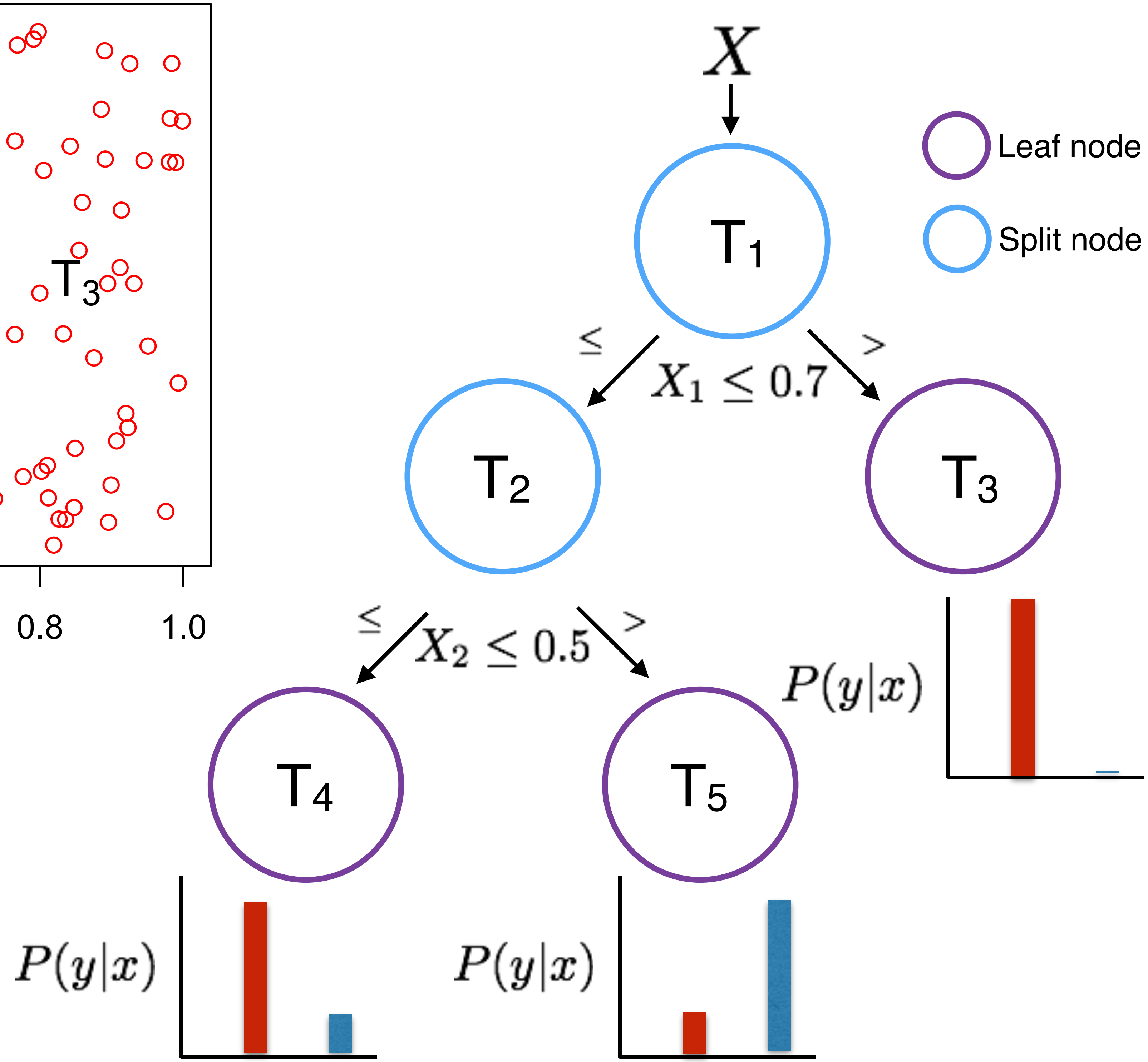
From decision trees to extra trees



Schrider and Kern (2016)

From decision trees to extra trees

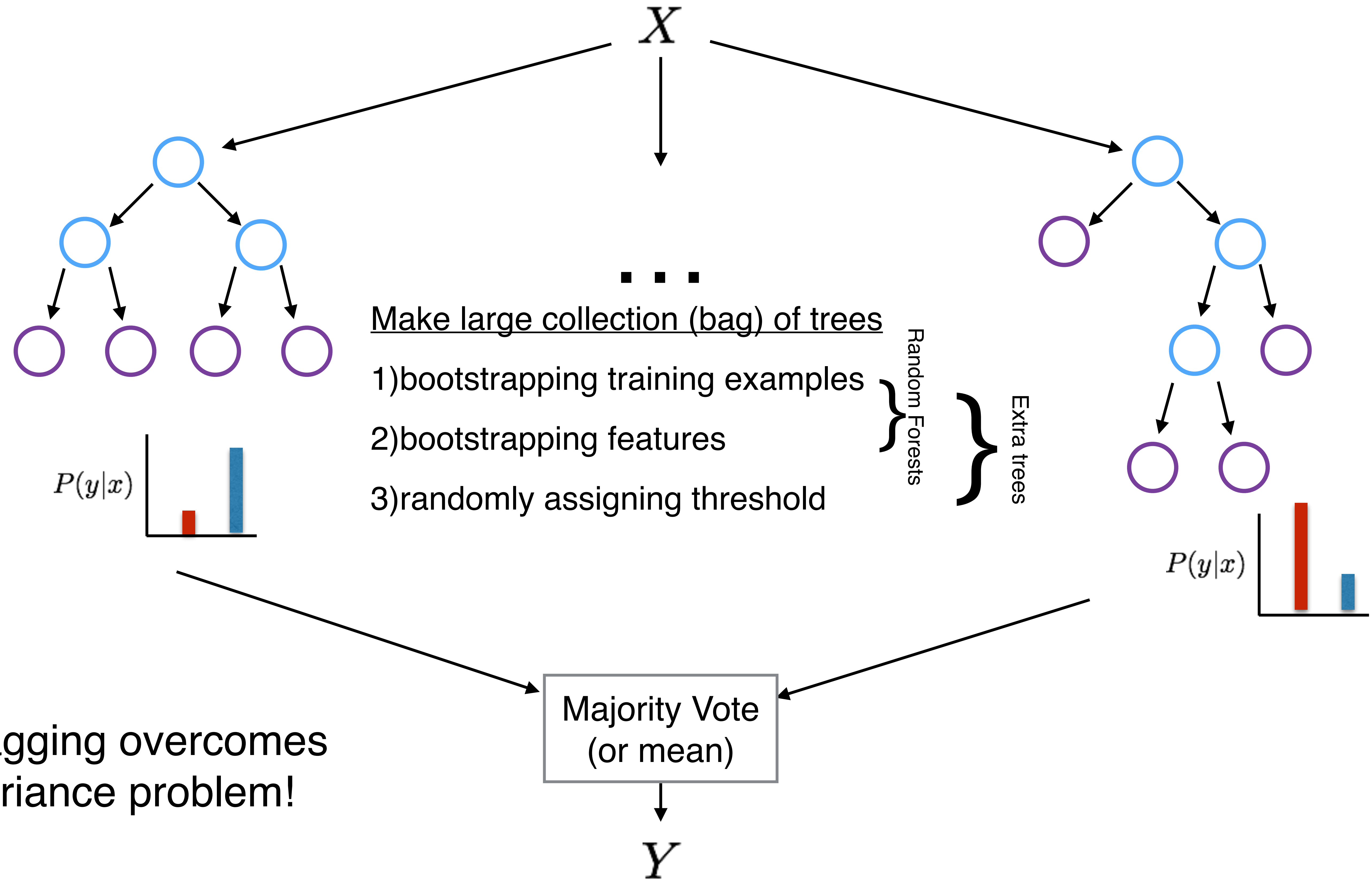




decision trees have low bias but suffer from high variance

Schrider and Kern (2016)

From decision trees to extra trees



bagging overcomes
variance problem!

Schrider and Kern (2016)