QUEEN'S UNIVERSITY BELFAST

SCHOOL OF ELECTRONICS, ELECTRICAL ENGNIEERING AND COMPUTER SCIENCE

Data Analysis & Visualisation

CSC3062

BEng (CS & SE), MEng (CS & SE), BIT & CIT

Dr Reza Rafiee

Semester 1 2019

Supervised learning

What we need to know about classification

- What is classification?
- What we need as a dataset in classification
- Binary vs. multiclass classification
- Classification models (categories of classifier models)
- How to choose a classification model?
- Support vector machine (SVM) classifier model
- Designing a multiclass SVM model with an example
- How to evaluate the performance of a classifier model?

Classification applications

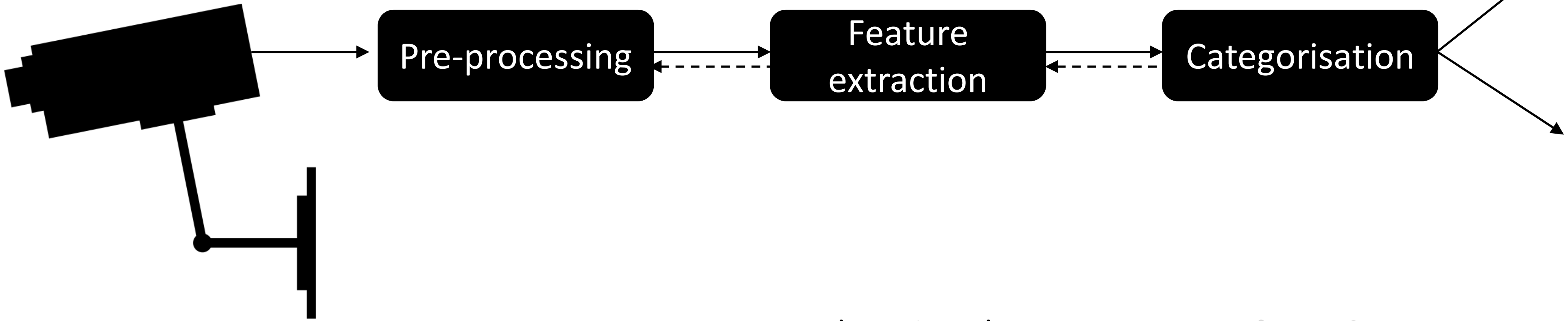
- Text categorisation (e.g., spam filtering)
- Optical character recognition (OCR) (e.g., a computerised system transferring hard documents into word doc.)
- Machine vision (e.g., face detection)
- Natural-language processing (NLP) (e.g., spoken language understanding)
- Market segmentation (e.g., predict if customer will respond to promotion)
- Bioinformatics and medicine (e.g., classify cancer patients into different immune subtypes)
- …

QUEEN'S UNIVERSITY BELFAST

SCHOOL OF ELECTRONICS, ELECTRICAL ENGNIEERING AND COMPUTER SCIENCE

Pattern recognition systems

Set up a camera (or transducer)



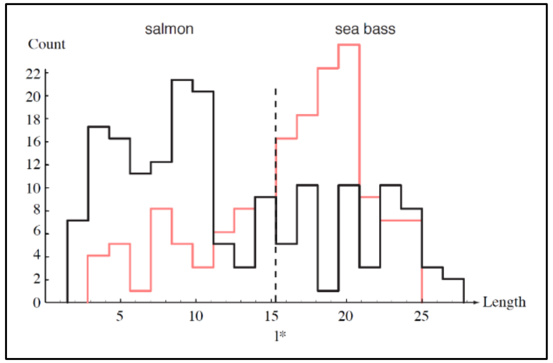
Object(s) (samples)

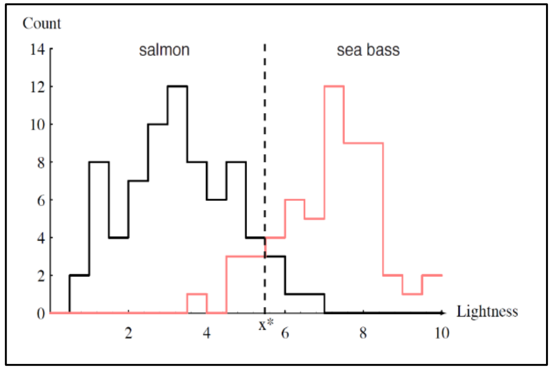
Measurement

Filtering, removing noise and so on

Choosing the best features

Clustering
Classification





x_1: lightness
x_2: width

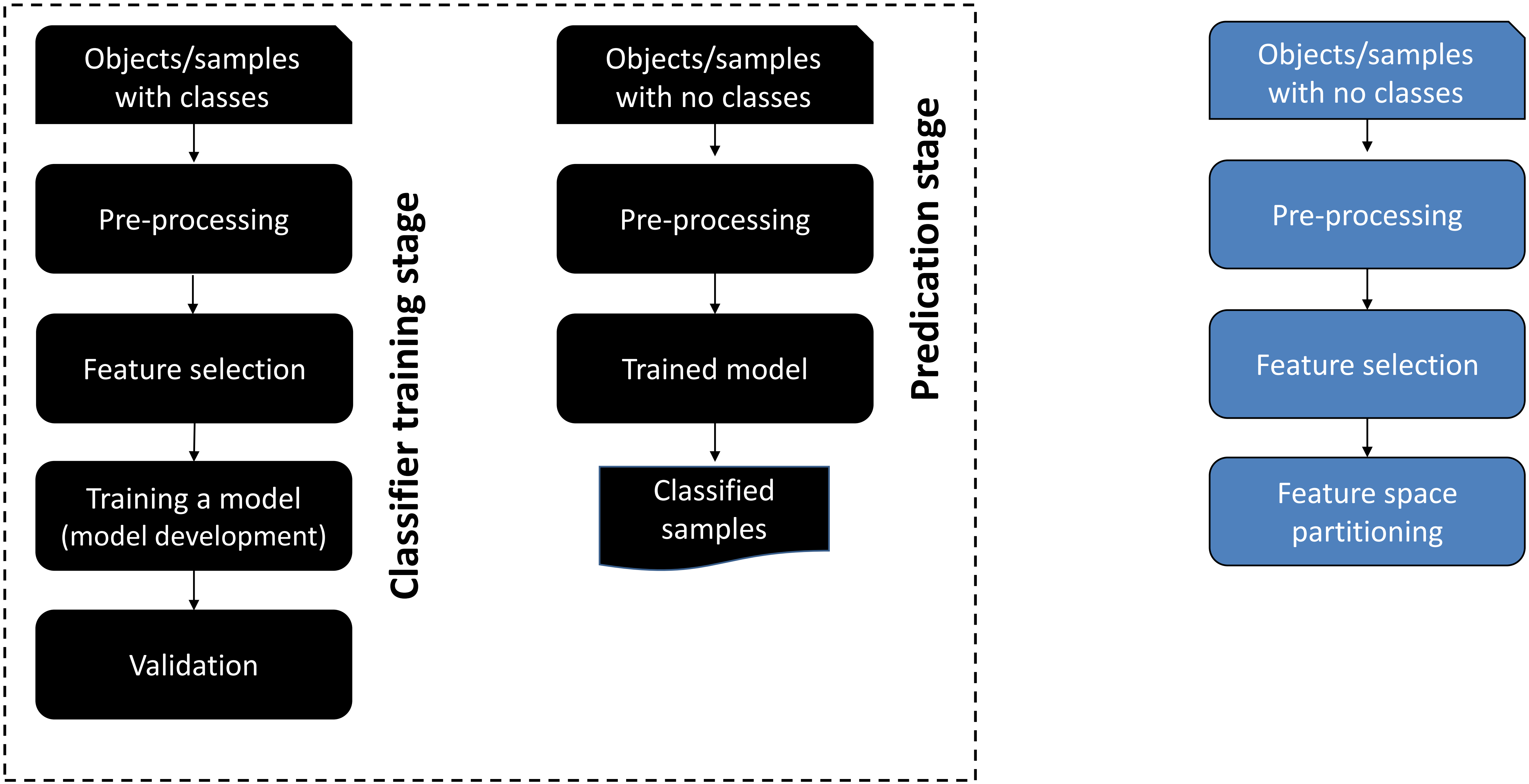
The aim is to partition the feature space into two regions

Feature space: two dimensions $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

Suppose that we measure the feature vectors for our samples

Reference: Pattern Classification, R. O. Duda, P. E. Hart, D. G. Stork; Wiley

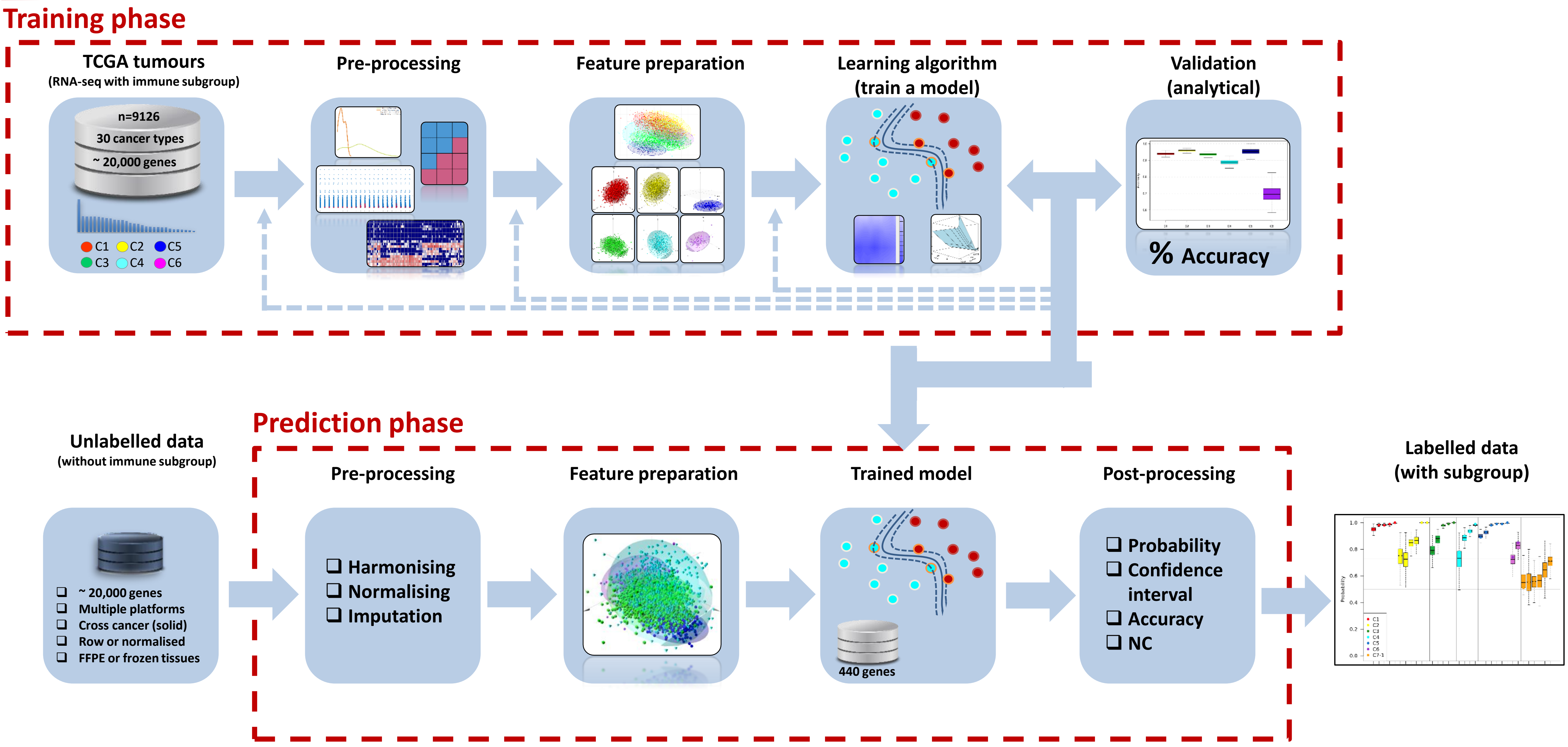
Classification vs. clustering



Feature selection for high separability

QUEEN'S UNIVERSITY BELFAST
SCHOOL OF ELECTRONICS, ELECTRICAL ENGNIEERING AND COMPUTER SCIENCE

Classification – training vs. prediction



Classification stages. a, Training phase. **b**, Prediction phase.

Prediction problem

Consider the following example

It is important for the bank to be able to **predict** in advance **the risk associated with a loan**, which is the probability that the customer will default and not pay the whole amount back.

In **credit scoring**, the bank calculates the risk given the amount of credit and **the information about the customer**. The **information about the customer** includes data we have access to and is relevant in calculating his or her financial capacity - namely, **income**, **savings**, **collaterals**, **profession**, **age**, **past financial history**, and so forth. The bank has a record of past loans containing such customer data and whether the loan was paid back or not.

From this data of particular applications, the aim is **to infer a general rule** coding **the association between a customer's attributes and his risk**.

A machine learning system fits a model to the past data to be able to calculate the risk for a new application and then decides to accept or refuse it accordingly.

Originated from "Introduction to Machine Learning", Ethem Alpaydin, MIT Press, 2004

Prediction problem - classification

- It is important for the bank to be able to **predict** in advance **the risk associated with a loan**, which is the probability that the customer will default and not pay the whole amount back.

- From a data of particular applications, the aim is to infer a general rule coding **association** between a **customer's attribute (features)** and **his/her risk**.

- This is an example of a ***classification*** problem where there are two classes: low-risk and high-risk customers. The information about a customer makes up the ***input*** to the classifier whose task is to assign the input to one of the two classes.

Dataset in classification

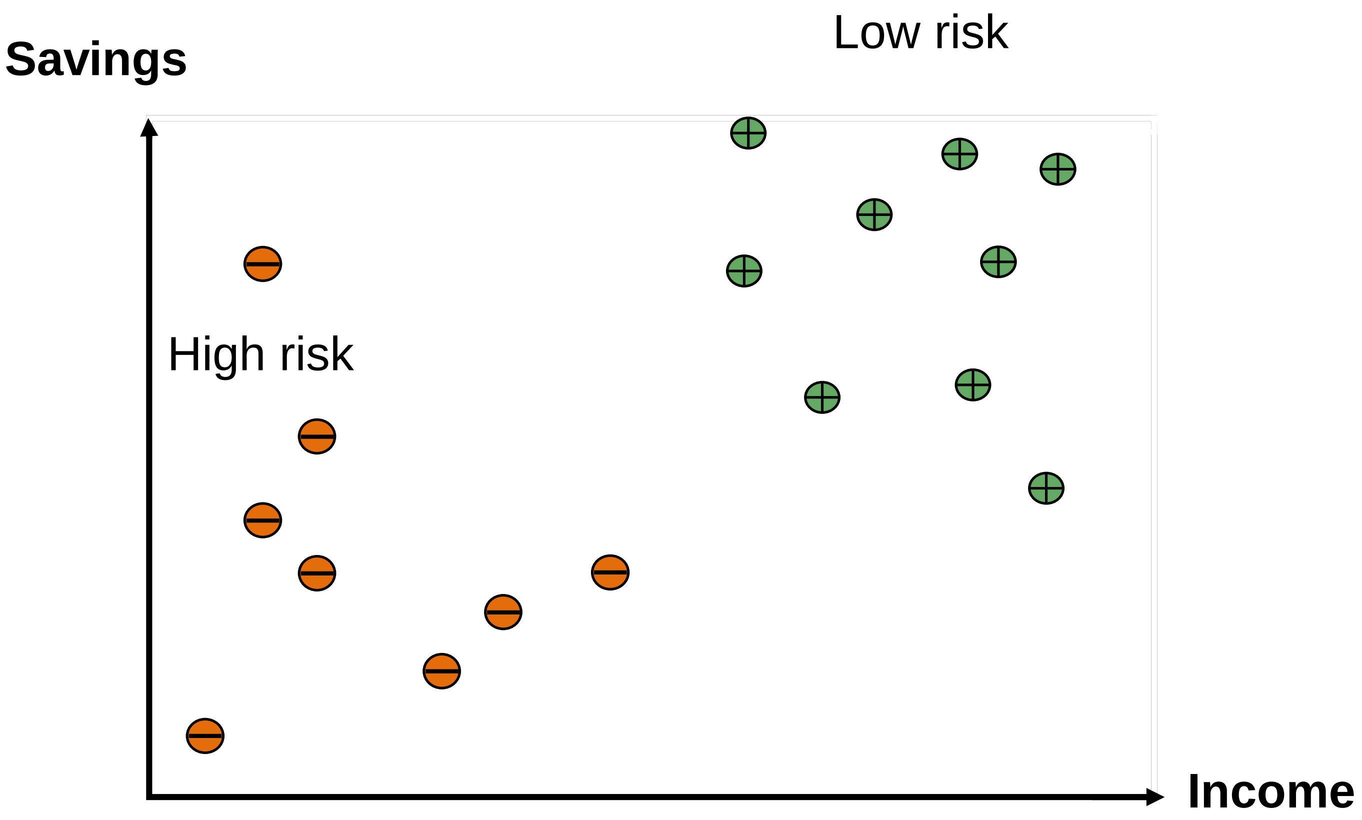
Features

17 samples (customers)

#	Income	Savings	Collaterals	Profession	age	Risk
1	£25K	2K	Yes	…	24	low
2	£35K	5K	No	…	30	low
3	£5K	0.5K	Yes	…	26	high
…	…	…	…	…	…	..
17	£45K	10K	No	…	40	low

Class labels or subgroups

Dataset in classification



#	Income	Savings	Collaterals	Profession	age	Risk
1	£25K	2K	Yes	...	24	low
2	£35K	5K	No	...	30	low
3	£5K	0.5K	Yes	...	26	high
...	...	...	...	...	...	..
17	£45K	10K	No	...	40	low

This figure illustrates an example of a dataset. Each circle corresponds to one **data instance** with input values in the corresponding axes. **For simplicity**, only two customer **attributes or features**, income and savings, are taken **as input** and the two classes are low-risk (‘+’) and high-risk (‘−’).

Prediction problem – *discriminant*

X_1: Income

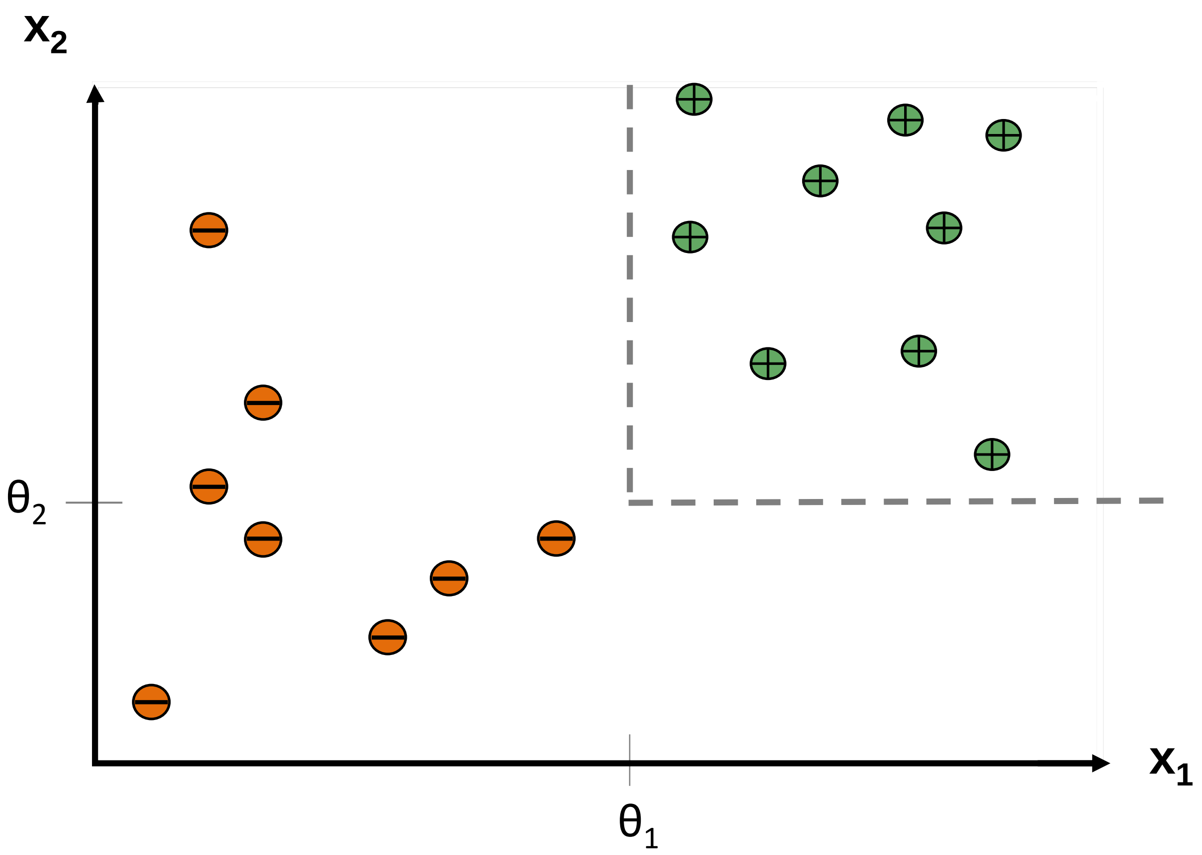
X_2: Savings

Y: low-risk or high-risk

- This is an example of *discriminant*; it is a function that separates the examples of different classes.
- **Discriminant analysis**

Prediction problem; parameters of the model

Model parameters: θ_1 & θ_2



IF $x_1 > \theta_1$ and $x_2 > \theta_2$ THEN low-risk ELSE high-risk

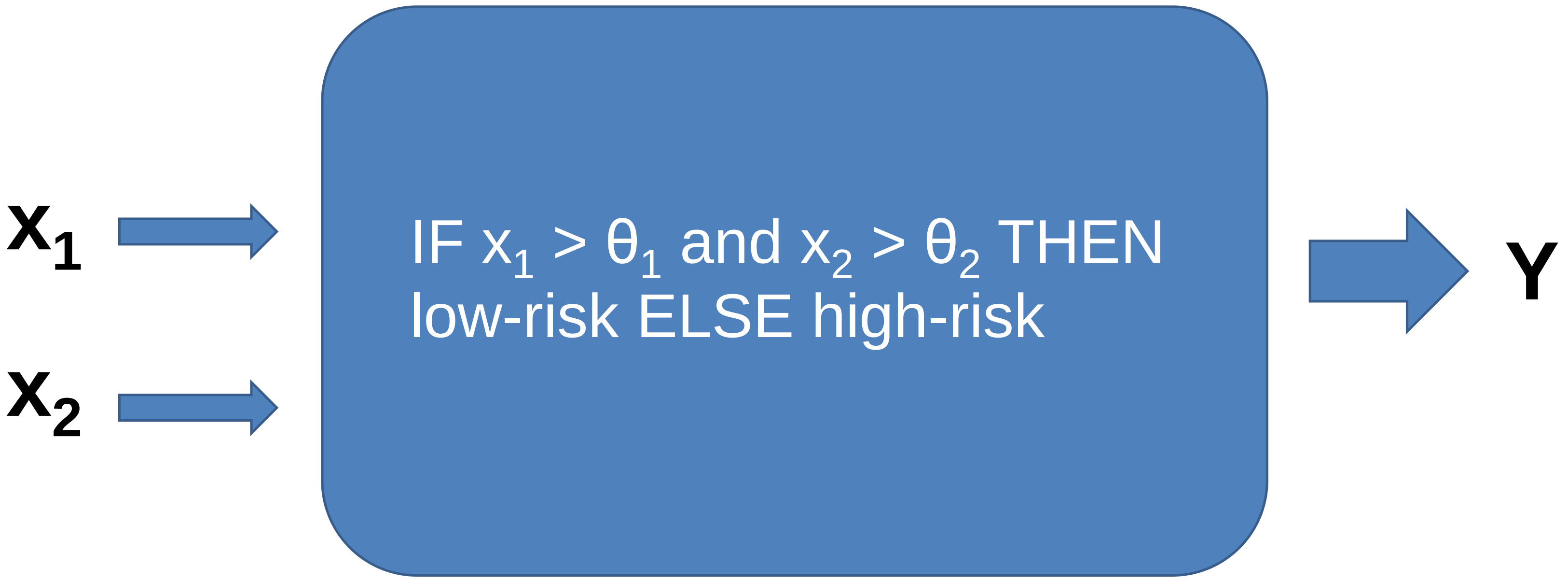
Once we have a rule like this that fits the past data, if the future is similar to past, then we can make correct predications for new instances.

Prediction problem – *discriminant*

X_1: Income

X_2: Savings

Y (output): low-risk or high-risk



- This is an example of *discriminant*; it is a function that separates the examples of different classes.
- Discriminant analysis

Classification - concept

Basic idea: classify samples into pre-defined groups

What is difference between classification and regression?

Regression vs. classification

Basic idea: classify samples into pre-defined groups

A regression algorithm predicts a continuous output

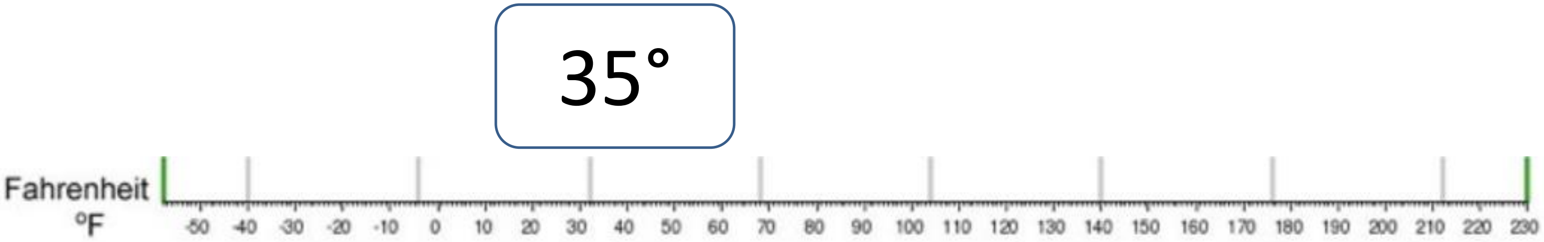
A classification algorithm predicts a discrete output (class label)

QUEEN'S UNIVERSITY BELFAST | SCHOOL OF ELECTRONICS, ELECTRICAL ENGNIEERING AND COMPUTER SCIENCE

Prediction in regression & classification

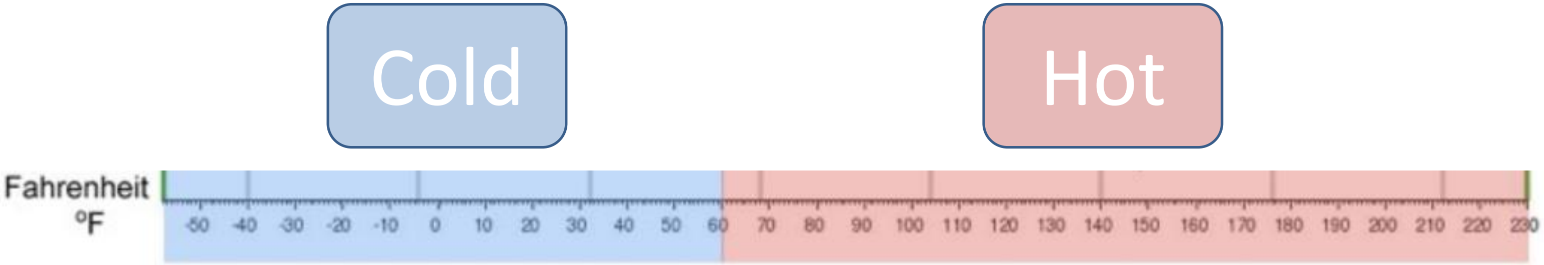
What is the temperature going to be tomorrow?

Based on previous data



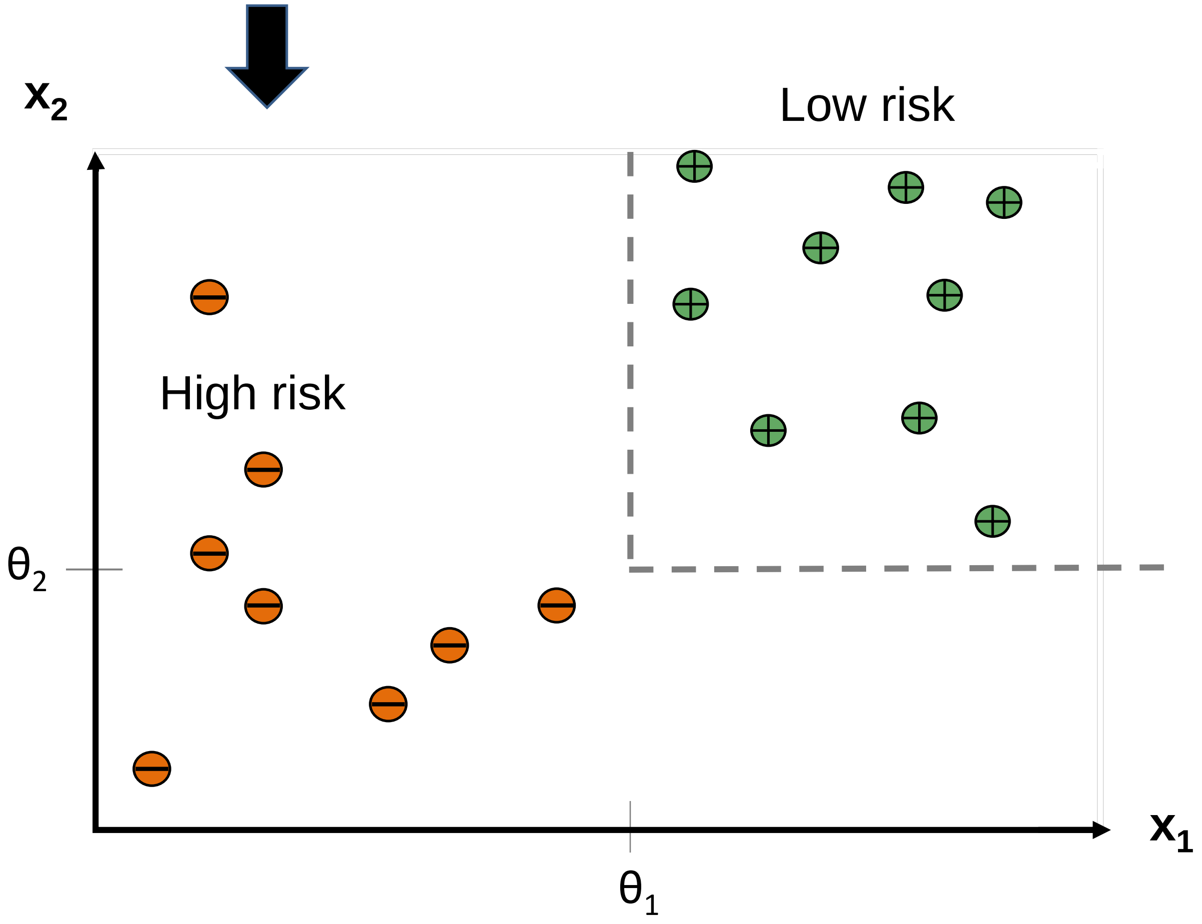
Will it be cold or hot weather?

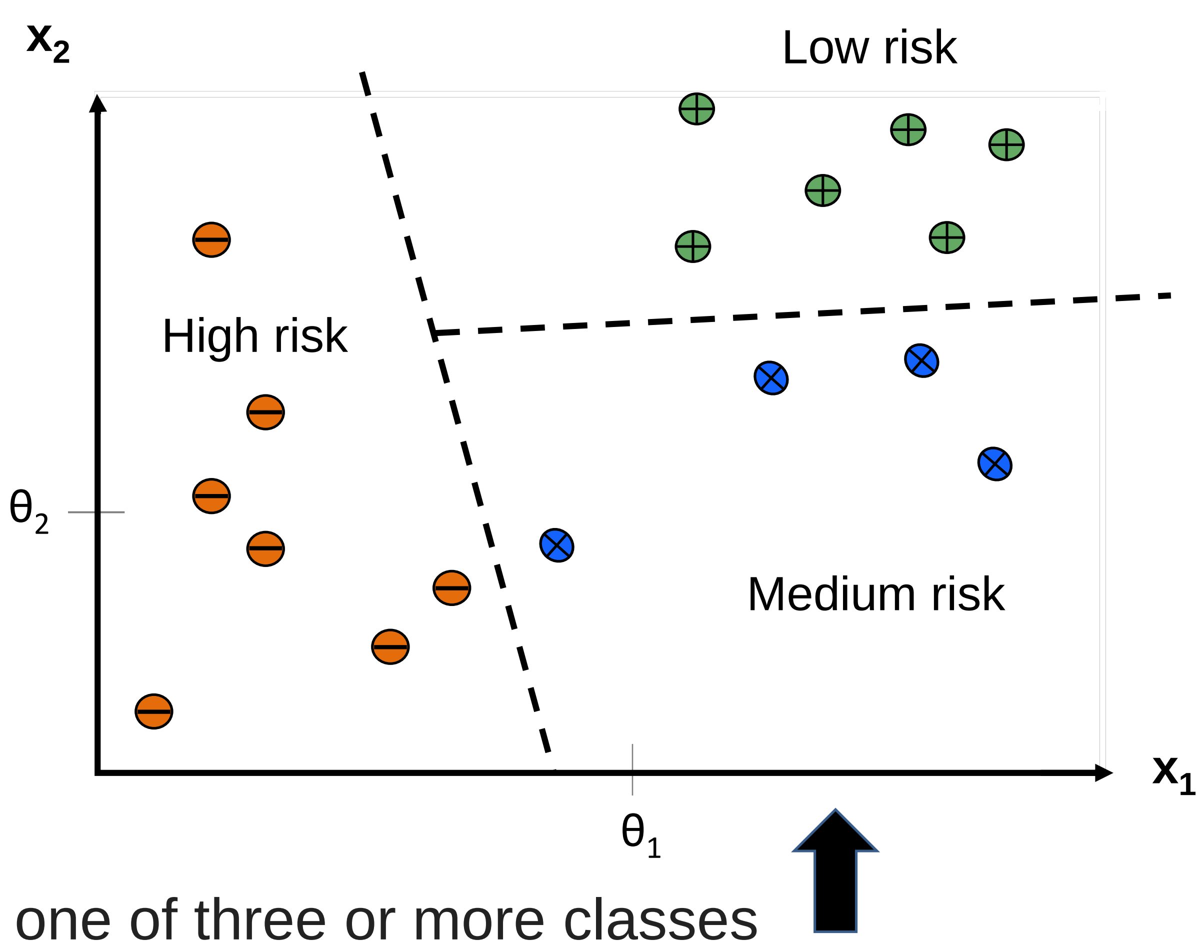
Based on previous data



Binary vs. multiclass classification

Binary classifier classifies data points into one of two classes





Multiclass classifier: classifies data points into one of three or more classes

Classification algorithms

- K-Nearest Neighbour
- Naive Bayes Classifier
- Support Vector Machines (the basic SVM supports only binary classification); linear or with Gaussian kernels
- Decision Trees (e.g., Random Forest)
- Artificial Neural Networks (ANN)
- Hierarchal classifier
- …

Classification algorithms

- ❑ K-Nearest Neighbour
- ❑ Naive Bayes Classifier
- ❑ Support Vector Machines (the basic SVM supports only binary classification); linear or with Gaussian kernels
- ❑ Decision Trees (e.g., Random Forest)
- ❑ Artificial Neural Networks (ANN)
- ❑ Hierarchal classifier
- ❑ …

Parametric vs. nonparametric models

In **parametric** “learners” or models, we have a fixed size of parameters (the weight coefficient). This makes them not very flexible.

- Linear Regression
- Linear Support Vector Machines
- Logistic Regression
- Naive Bayes

Parametric vs. nonparametric models

While **nonparametric** “learners” or models are good when you have a large data and you don’t want to worry too much about choosing just the right features

- ❑ Decision Trees
- ❑ K-Nearest Neighbour
- ❑ Support Vector Machines with Gaussian Kernels
- ❑ Artificial Neural Networks

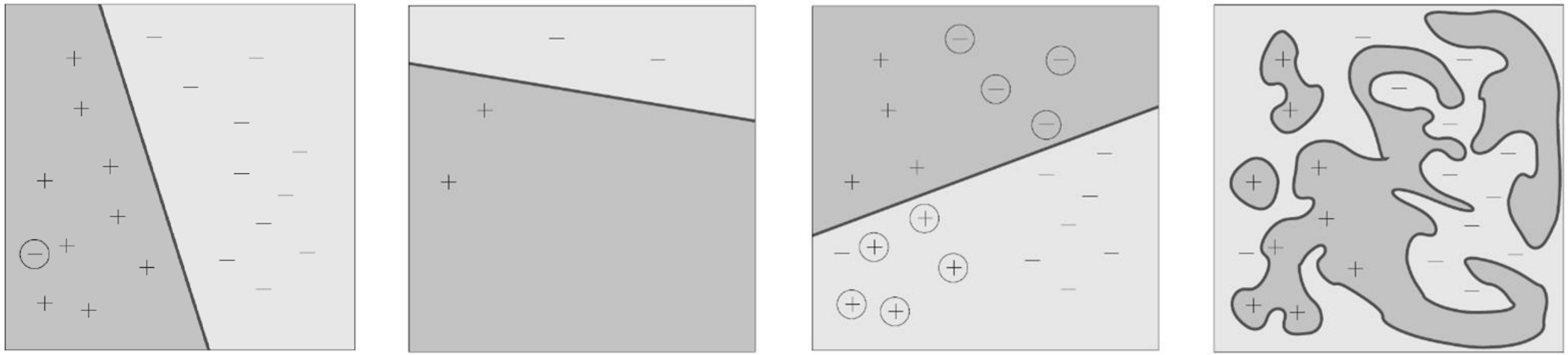
Non-parametric models assume that the data distribution cannot be defined in terms of such a finite set of parameters. This makes them more flexible

Parametric vs. nonparametric models

Parametric	Nonparametric
Linear regression	Decision Trees
Naive Bayes	KNN
Linear SVMs	SVMs (nonlinear kernels)
Logistic regression	ANNs
Less flexibility	More flexibility

Good vs. bad classifiers?

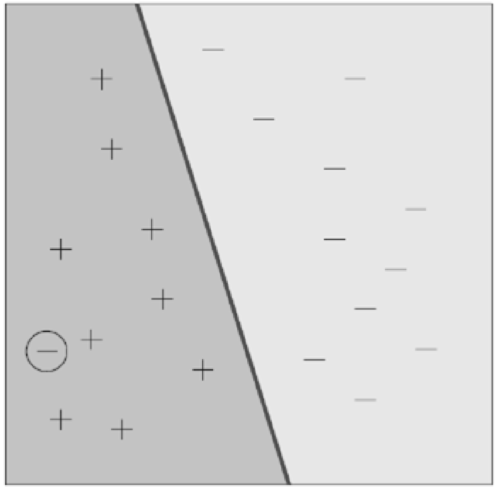
Assume, we have four classifiers developed on four training datasets.



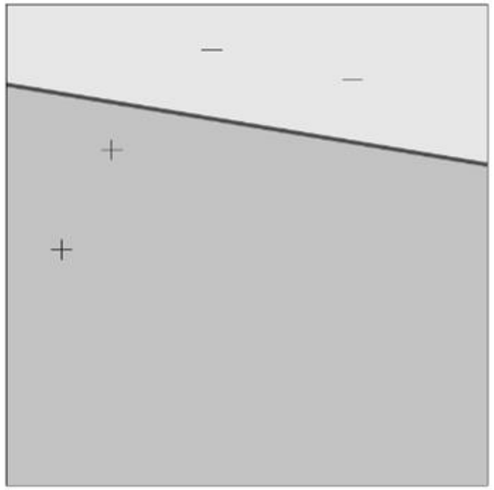
Good vs. bad classifiers

Good

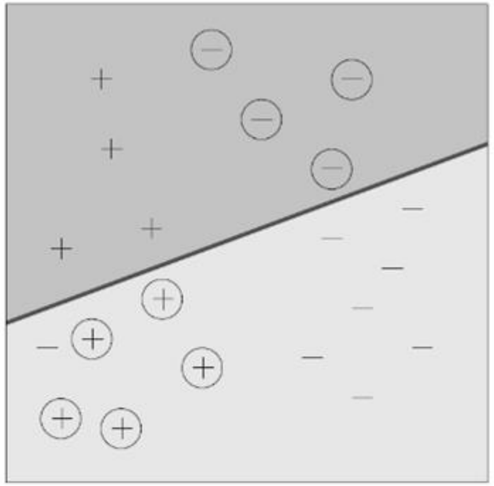
Bad



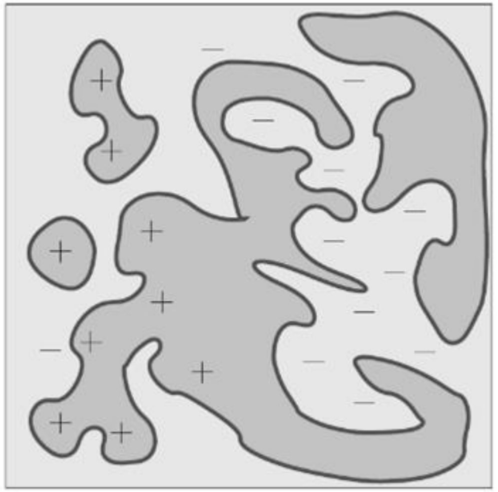
Sufficient data
Low training error
Simple classifier



Insufficient data



Training error too high



Classifier too complex

Any Questions?