

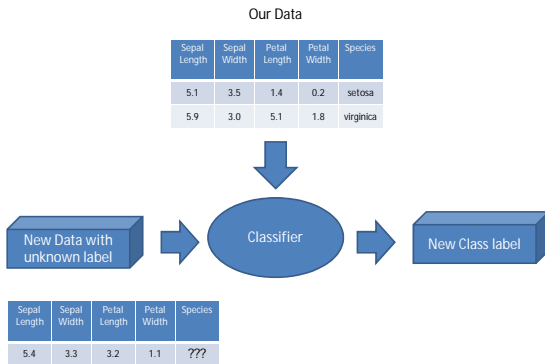
# **Introduction to Machine Learning**

## **Classification: Tasks**

[compstat-lmu.github.io/lecture\\_i2ml](https://compstat-lmu.github.io/lecture_i2ml)

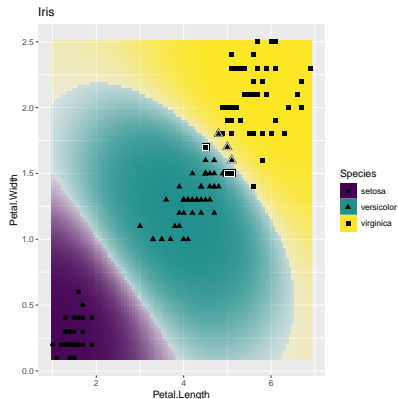
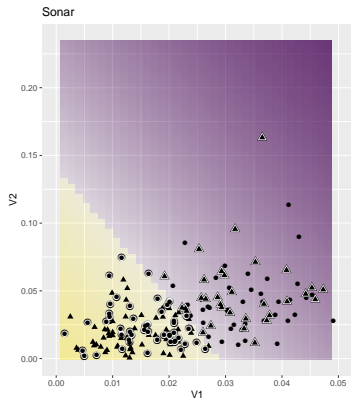
# CLASSIFICATION

Learn functions that assign class labels to observation / feature vectors. Each observation belongs to exactly one class. The main difference to regression is the scale of the output / label.



# BINARY AND MULTICLASS TASKS

The task can contain 2 classes (binary) or multiple (multiclass).



# BINARY CLASSIFICATION TASK - EXAMPLES

- Credit risk prediction, based on personal data and transactions
- Spam detection based on textual features
- Churn prediction based on customer behavior
- Predisposition for specific illness based on genetic data

## Do polygraphs detect lies?

*Polygraph or "lie detector" exams continue to be used by law enforcement and government agencies for various screenings even though most criminal courts ban polygraph evidence.*

### How reliable?

**Supporters** claim an 85-95 percent accuracy rate

**Critics** say there is not enough scientific evidence to say whether it detects lies or not

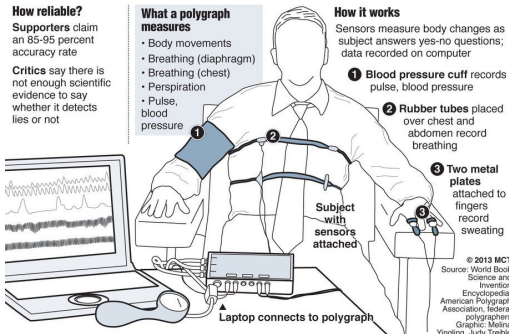
### What a polygraph measures

- Body movements
- Breathing (diaphragm)
- Breathing (chest)
- Perspiration
- Pulse, blood pressure

### How it works

Sensors measure body changes as subject answers yes-no questions; data recorded on computer

- 1 **Blood pressure cuff** records pulse, blood pressure
- 2 **Rubber tubes** placed over chest and abdomen record breathing
- 3 **Two metal plates** attached to fingers record sweating



© 2013 MCT  
Source: World Book  
Science and  
Invention  
Encyclopedia,  
American Polygraph  
Association, federal  
polygraphers  
Graphic: Melina  
Yingling, Judy Treble

# BINARY CLASSIFICATION TASK - EXAMPLES

<https://www.bendbulletin.com/localstate/deschutescounty/3430324-151/fact-or-fiction-polygraphs-just-an-investigative-tool>

# MULTICLASS TASK - MEDICAL DIAGNOSIS

INFO

SYMPTOMS

QUESTIONS

CONDITIONS

DETAILS

TREATMENT

Conditions that match your symptoms

UNDERSTANDING YOUR RESULTS 1

Acute Sinusitis

Moderate match

>

Influenza (flu) adults

Moderate match

>

Common cold

Fair match

>

Asthma (Teen and Adult)

Fair match

>

Whooping cough

Fair match

>

↓

LOAD MORE CONDITIONS

Gender **Female** Age **25** Edit

My Symptoms Edit  
**cough, headache, nausea**

Could you be pregnant? Edit  
**No**

↻

Start Over

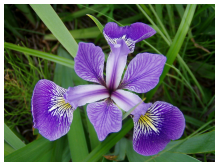
<https://symptoms.webmd.com>

# MULTICLASS TASK - IRIS

The iris dataset was introduced by the statistician Ronald Fisher and is one of the most frequent used datasets. Originally it was designed for linear discriminant analysis.



Setosa



Versicolor



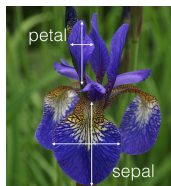
Virginica

Source:

[https://en.wikipedia.org/wiki/Iris\\_flower\\_data\\_set](https://en.wikipedia.org/wiki/Iris_flower_data_set)

# MULTICLASS TASK - IRIS

- 150 iris flowers
- Predict subspecies
- Based on sepal and petal length / width in [cm]



##		Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
##	1:	5.1	3.5	1.4	0.2	setosa
##	2:	4.9	3.0	1.4	0.2	setosa
##	3:	4.7	3.2	1.3	0.2	setosa
##	4:	4.6	3.1	1.5	0.2	setosa
##	5:	5.0	3.6	1.4	0.2	setosa
##	---					
##	146:	6.7	3.0	5.2	2.3	virginica
##	147:	6.3	2.5	5.0	1.9	virginica
##	148:	6.5	3.0	5.2	2.0	virginica
##	149:	6.2	3.4	5.4	2.3	virginica
##	150:	5.9	3.0	5.1	1.8	virginica



# MULTICLASS TASK - IRIS

