האוניברסיטה
העברית
בירושלים
THE HEBREW
UNIVERSITY
OF JERUSALEM

Artificial Intelligence in Medicine

Classification

Nir Friedman and Tommy Kaplan 8/1/24

Course goals

- Understand basic concepts in AI and ML
- Demystification

- How AI can help us?
- Formulate medical decisions as AI tasks

- How AI might fail us?
- Critical thinking - how to be a smart, responsible user, aware of system limitations

Administration

- 4 exercises (two exemptions) - 30% of grade
- Group project: AI in medicine idea/application
  20% of grade (in small groups)
- Exam - 50% of grade
- Quizzes (one exemption)
  10% magen: **max**(grade, 0.9*grade + 0.1*quizzes)

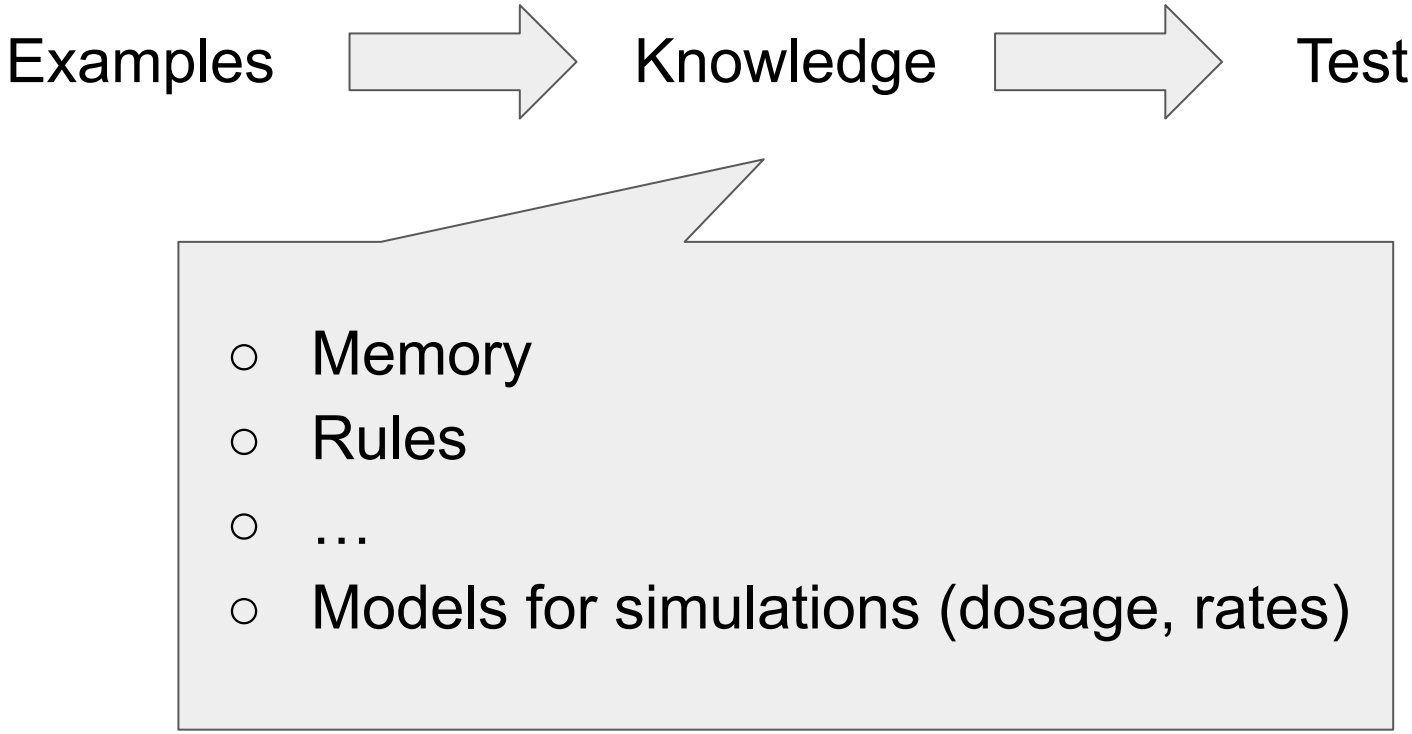
Prof. Tommy Kaplan (tommy@cs.huji.ac.il)

Prof. Nir Friedman (nir.friedman@mail.huji.ac.il)

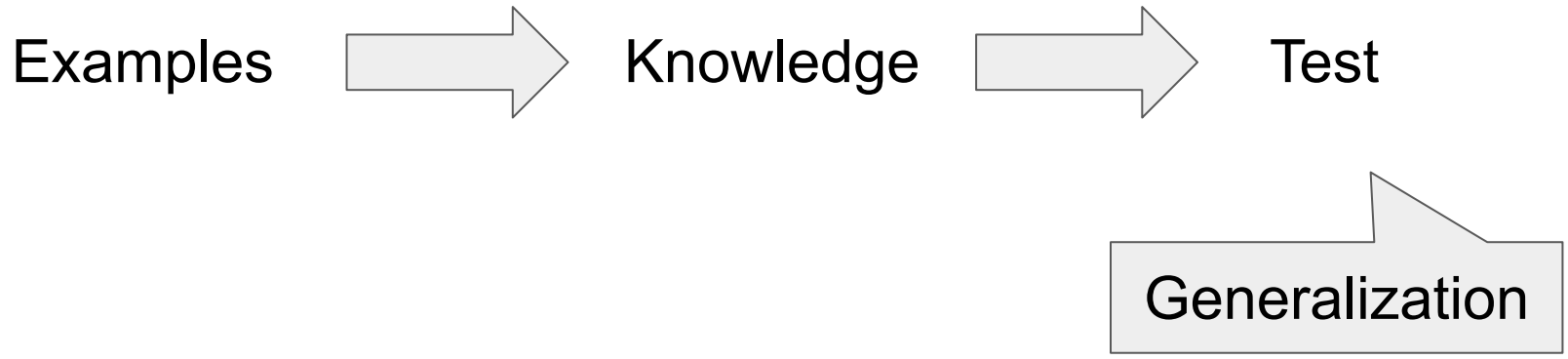
Roey Ben-Yosef (roey.benyosef@mail.huji.ac.il)

What is learning?

What is learning?



What is learning?



- Examples
  - memoization?
  - semantic search
  - clustering
  - predictions (smart elevators, waze timing)

What is machine learning?

Learning from data

- Understanding (how knowledge is represented)
  - Diagnostic rules
  - Cohort-based
  - Scenarios (predictions using known mechanisms)
  - Models for simulations (dosage, rates)

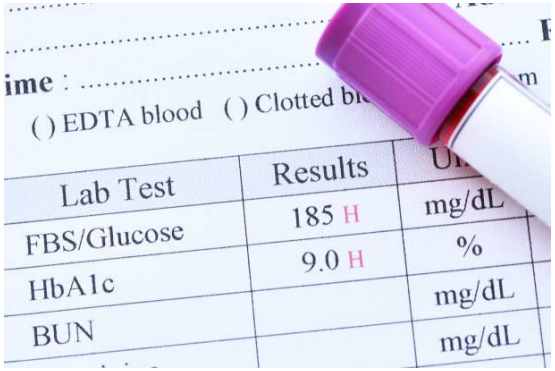
What is machine learning?

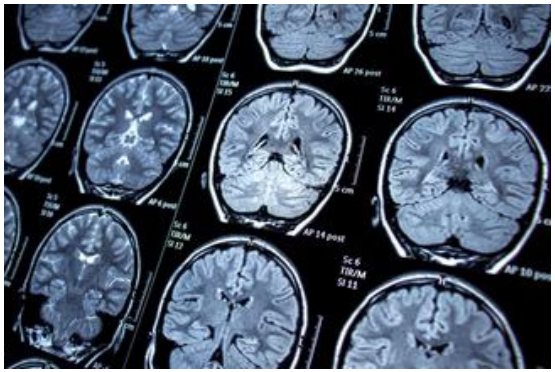
Learning from data

- Generalization
  - Is memoization learning?
  - semantic search
  - clustering
  - predictions (smart elevators, waze timing)

Medicine is data intensive

- Measurements, blood workup, imaging
- Digital devices and electronic records
- New sensors (longitudinal, wearable, internal)
- Millions of patients





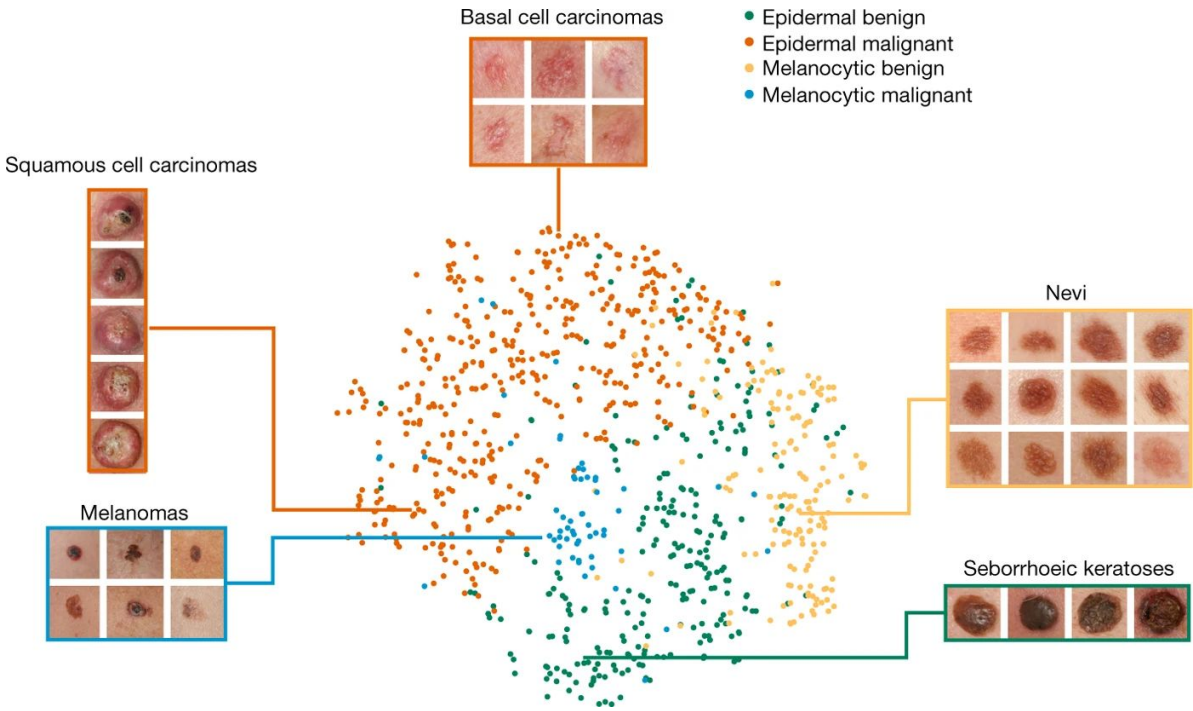
Syllabus

1	1/1	AI in ophthalmology (Prof. Itay Chowers)
2	8/1	Classification
3	15/1	Learning 1
4	22/1	Learning 2
5	7/2	Regression (Wed.)
6	12/2	Deep learning in image analysis (Prof. Leo Joskowicz)
7	19/2	Clustering
8	26/2	Dimensionality reduction and visualization
9	28/2	Deep learning, Missing data (Wed.)
10	4/3	Natural language in medicine (Dr. Gabi Stanovsky)
11	11/3	?

Classification

Fracture?

YES **NO**



https://www.nature.com/articles/nature21056

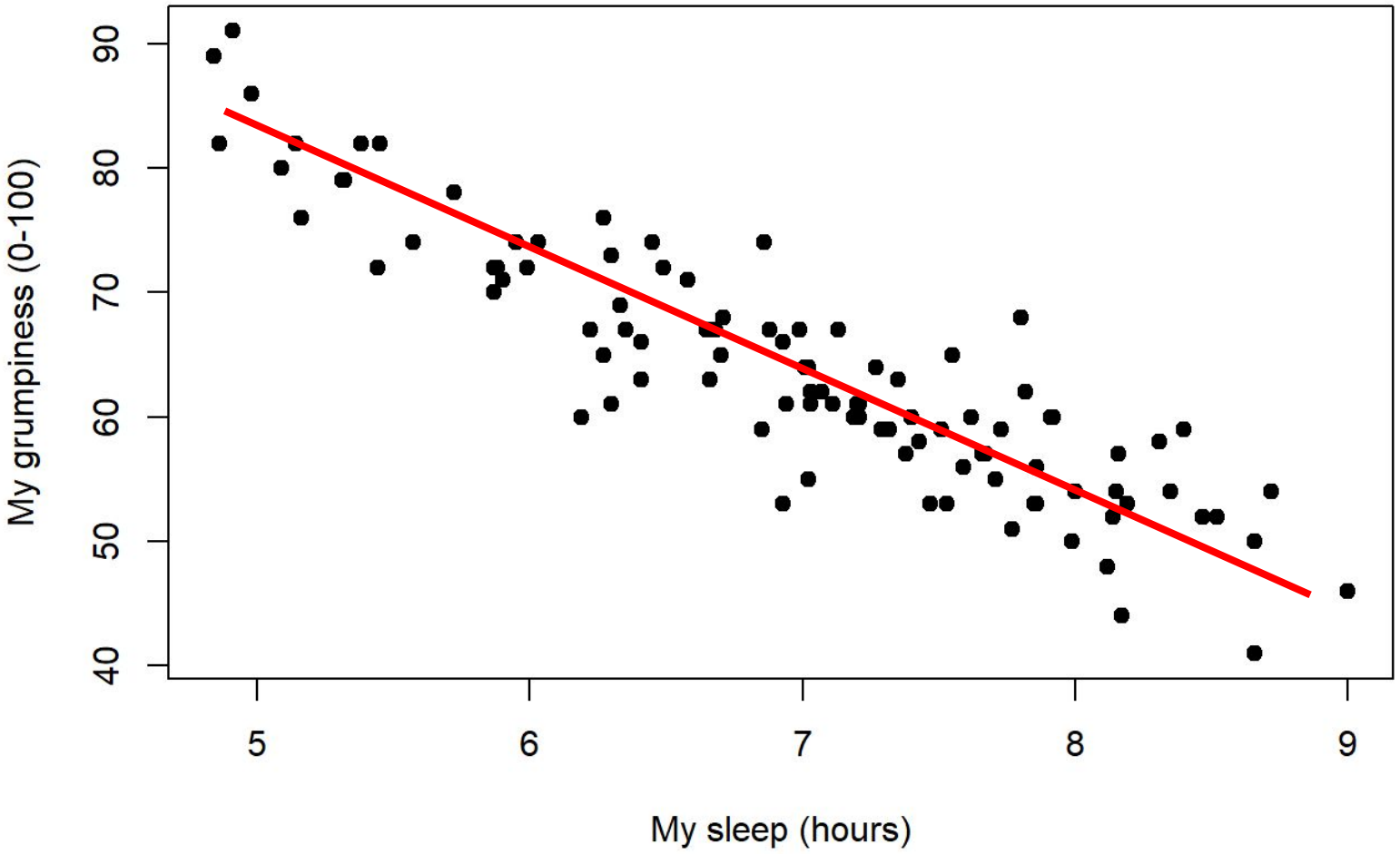
Learning?

Model selection?

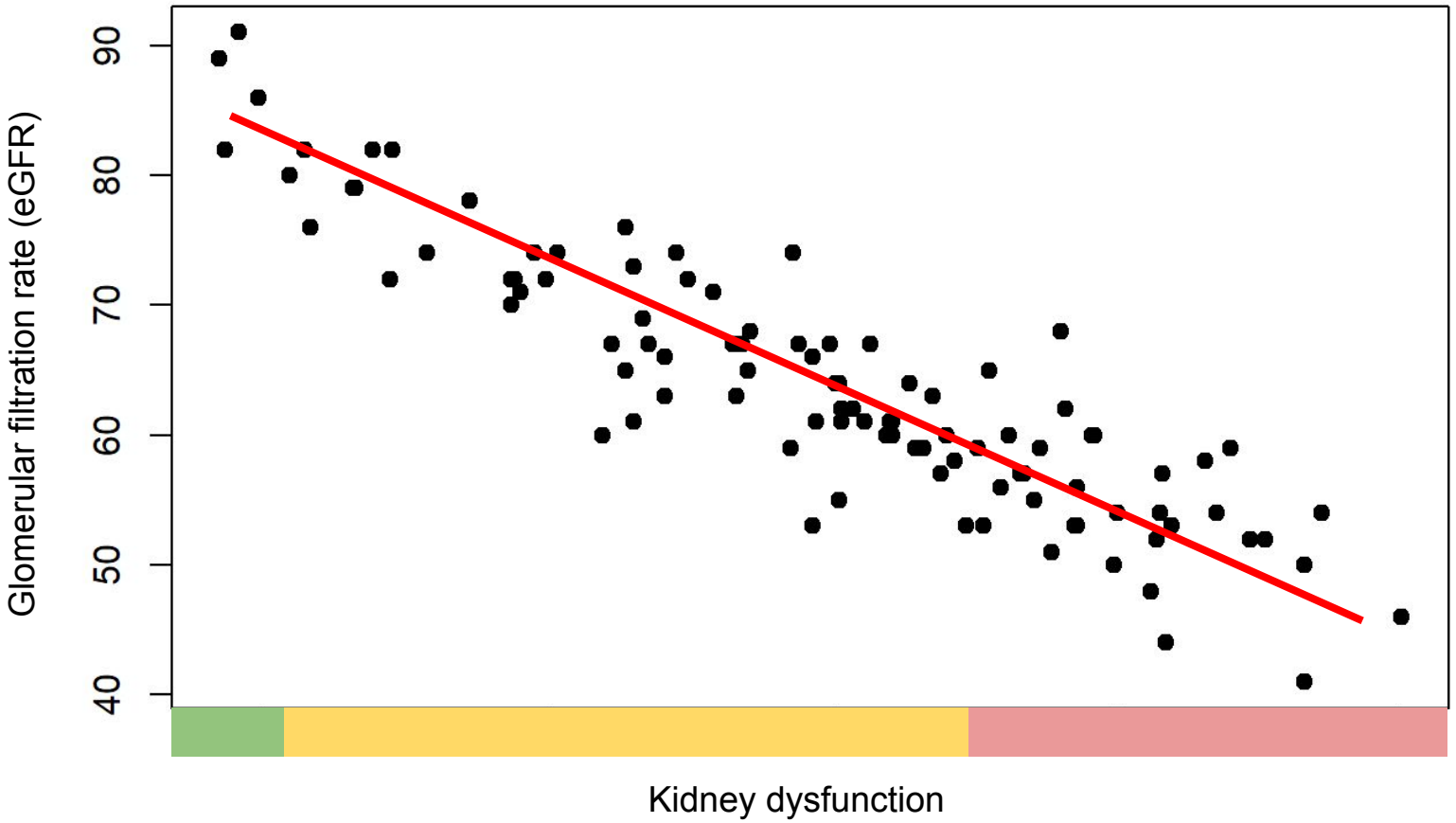
Classifier types

Over-fitting?

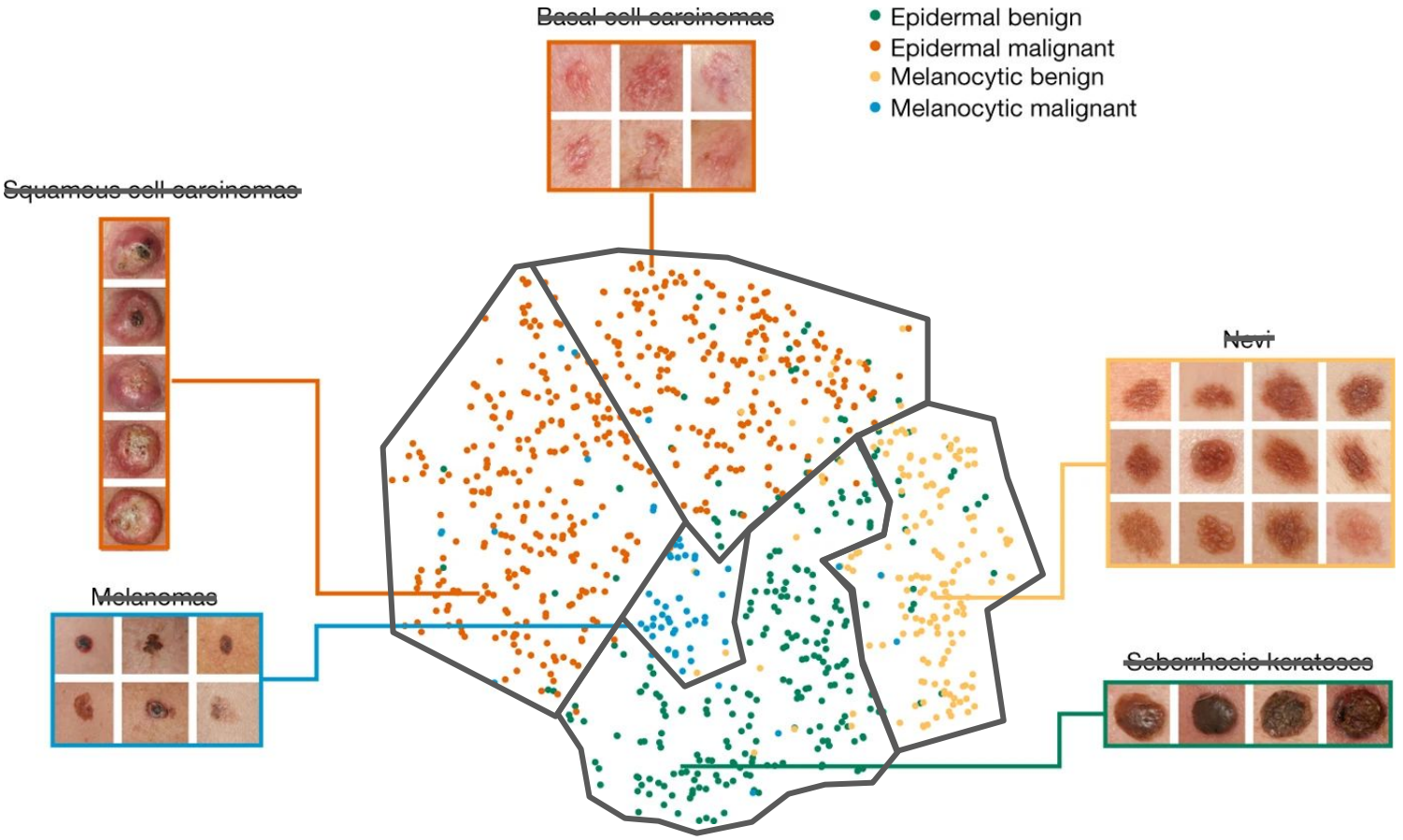
Regression



Regression

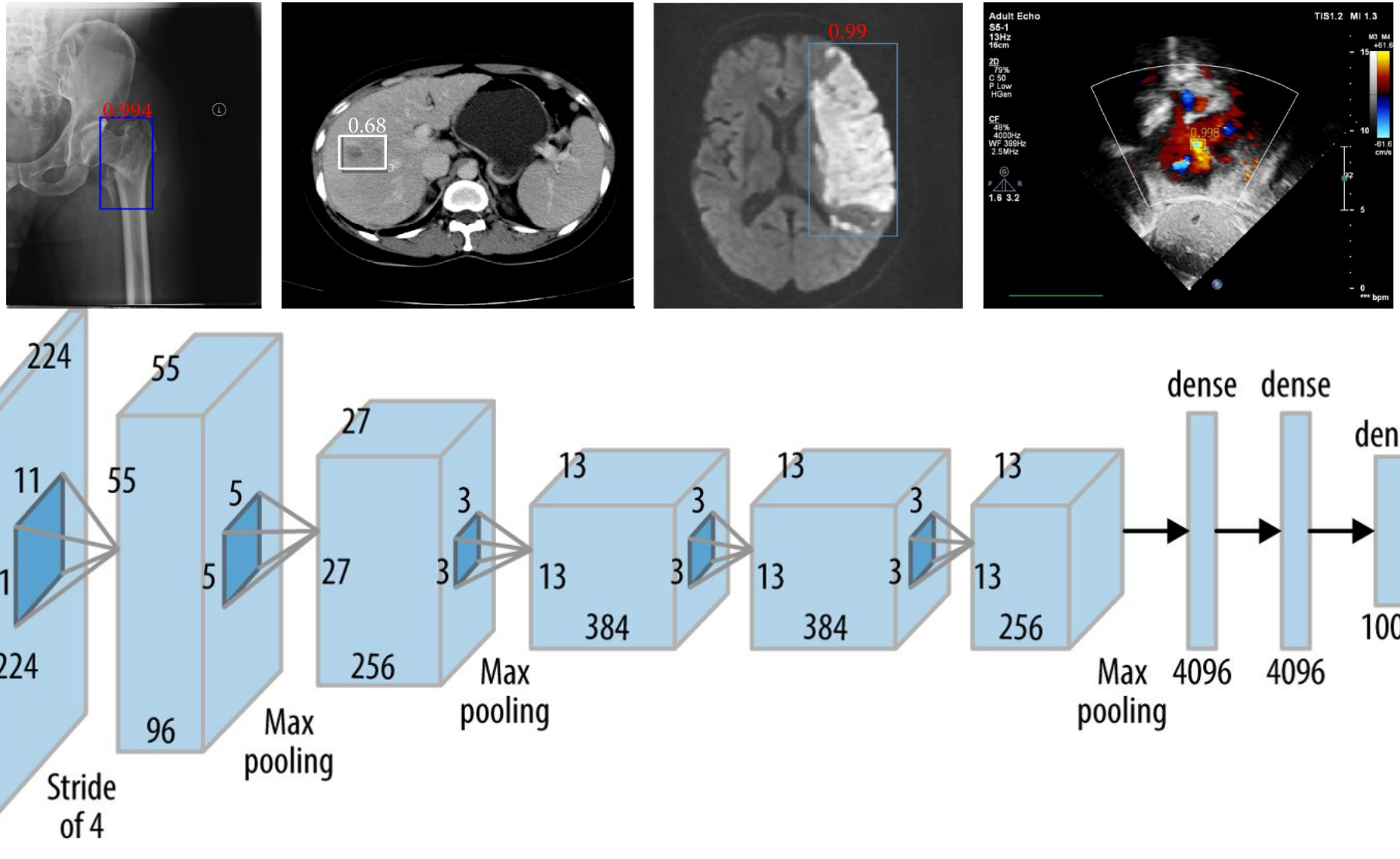

Glomerular filtration rate (eGFR)
90
80
70
60
50
40
Kidney dysfunction

Clustering



https://www.nature.com/articles/nature21056

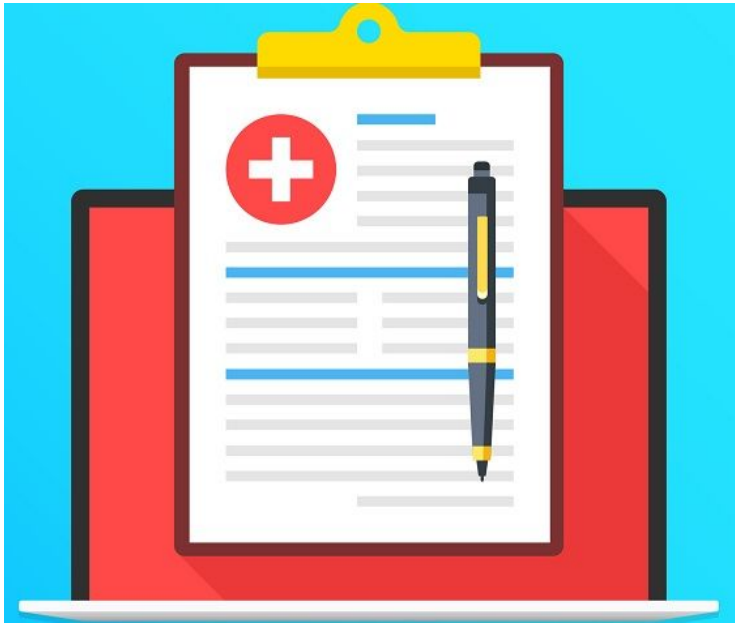
Deep learning in image analysis



https://spj.sciencemag.org/journals/hds/2021/8786793/ https://slazebni.cs.illinois.edu/spring22/lec01_intro.pdf

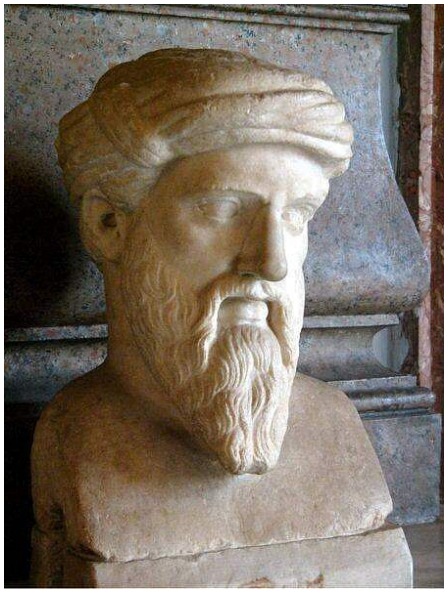
Natural language in medicine

- Clinical documentation
- Electronic health records (EHR)
- Text Classification
  - OCR
  - Tokenization
  - Lemmatization
  - Concept mapping
  - Topic modeling

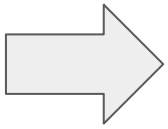


“The oldest, shortest words - 'yes' and 'no' - are those which require the most thought.”

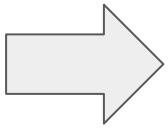
Pythagoras



Classifiers

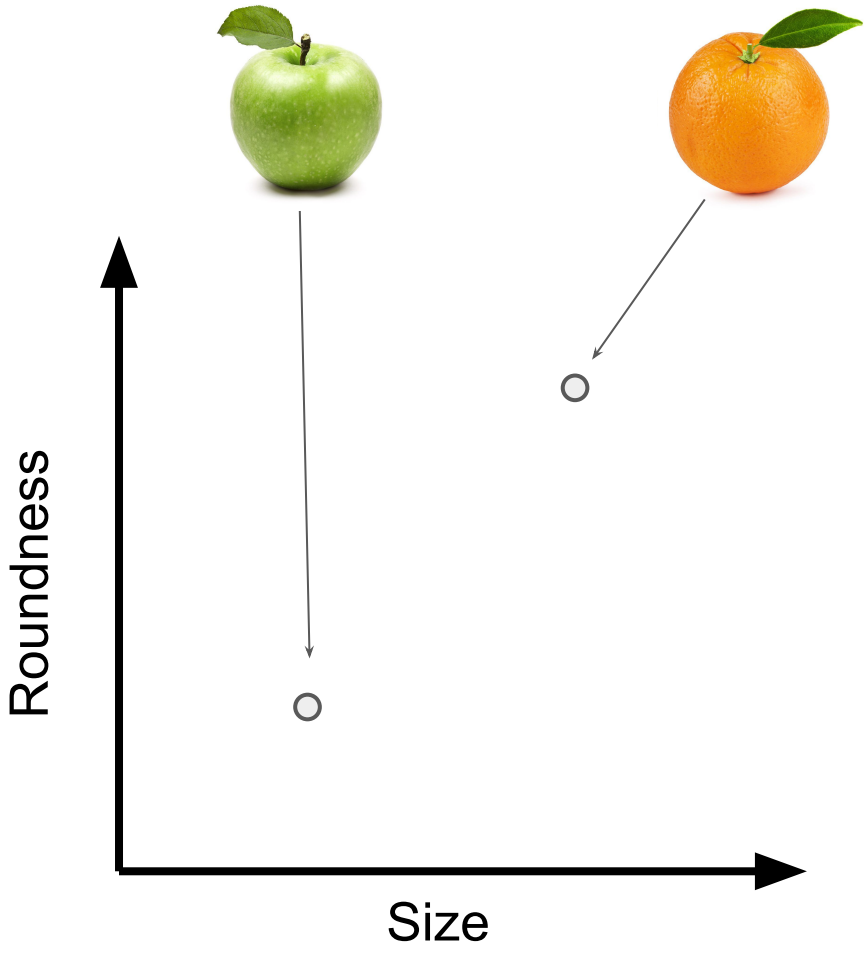


Classifier

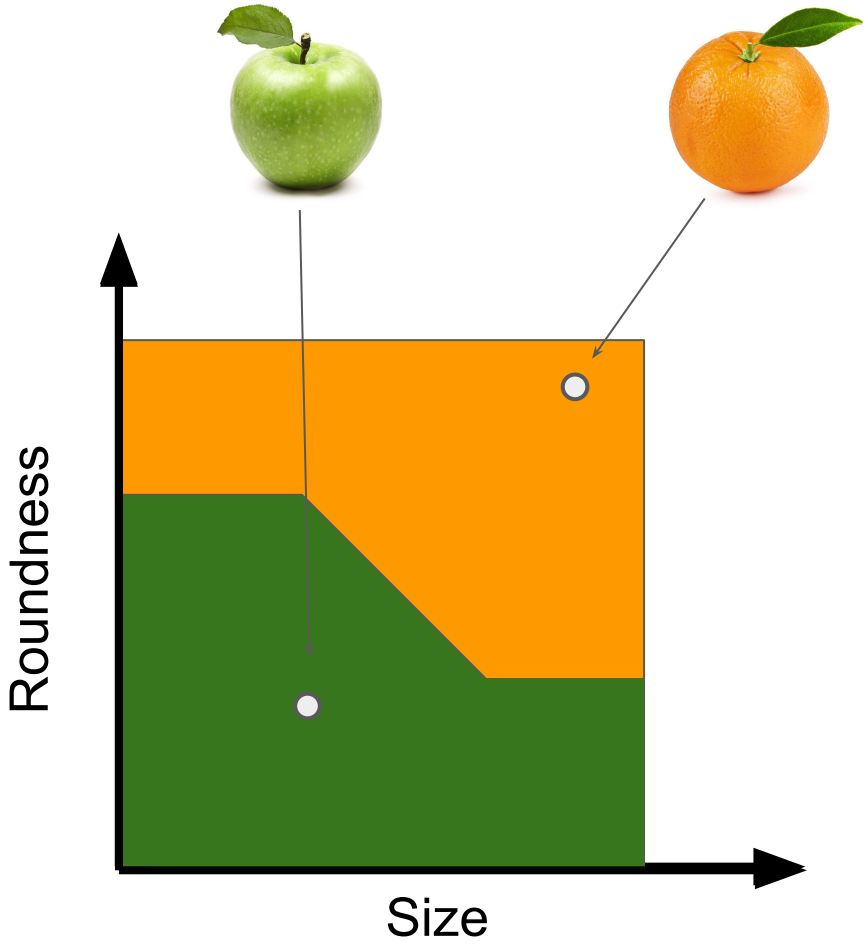


Apple
or
Orange

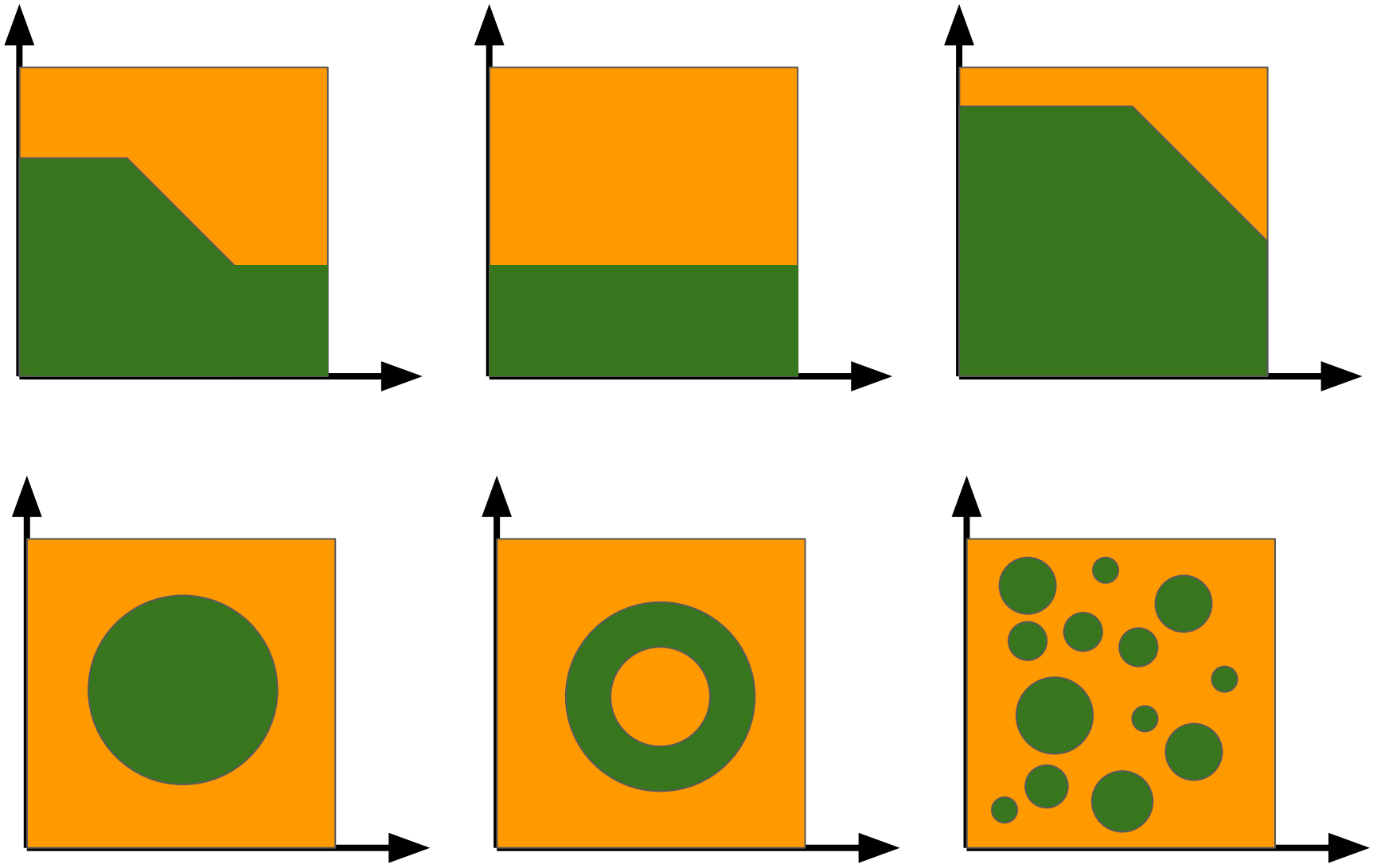
Classification Boundary



Classification Boundary



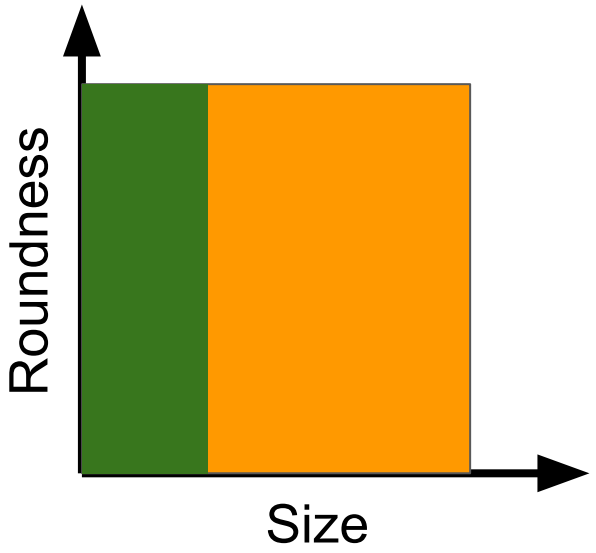
Hypotheses



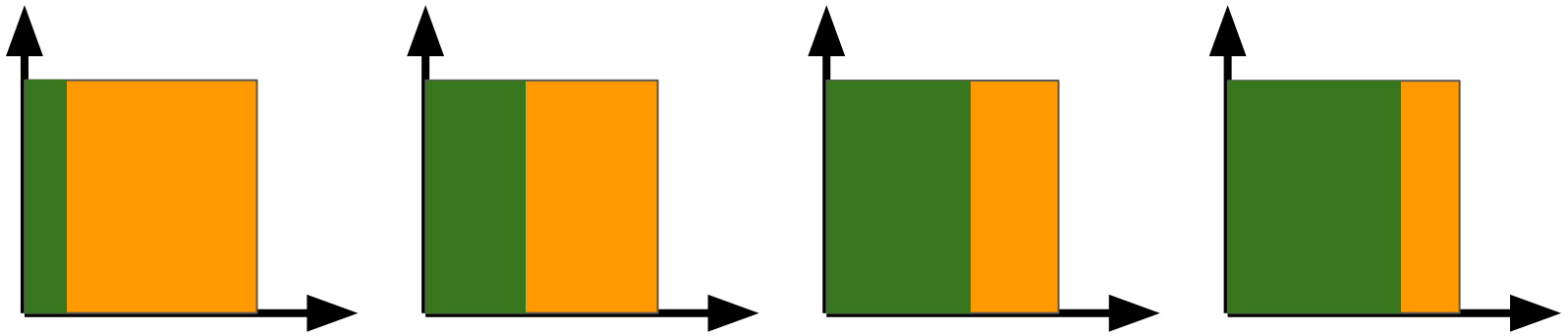
Representation

How to implement a classification?

```
if size > 100 then
   “orange”
else
   “apple”
```

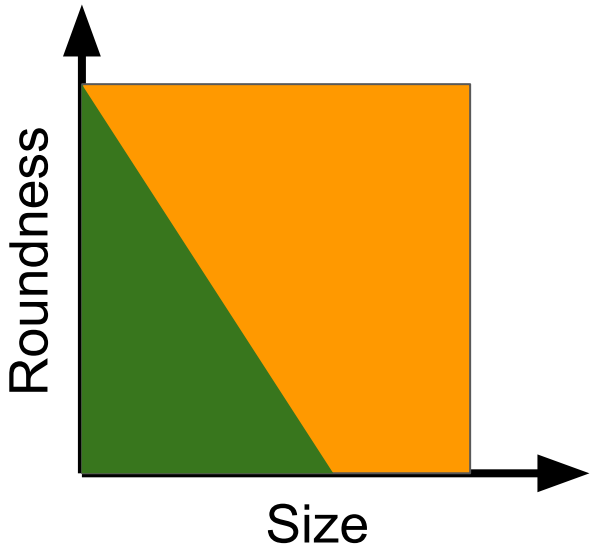


Whole range of hypotheses

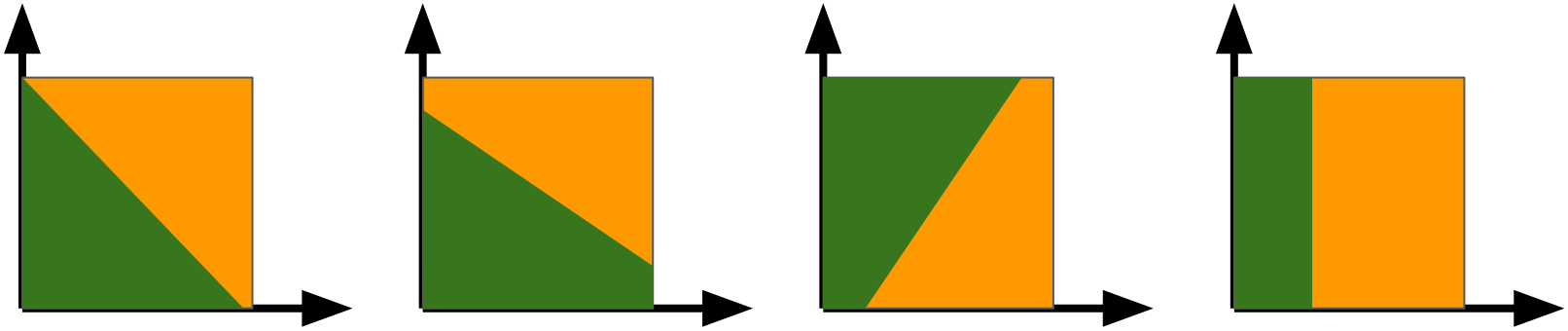


Linear classifier

```
if 2*size + 5*round > 100 then
   “orange”
else
   “apple”
```

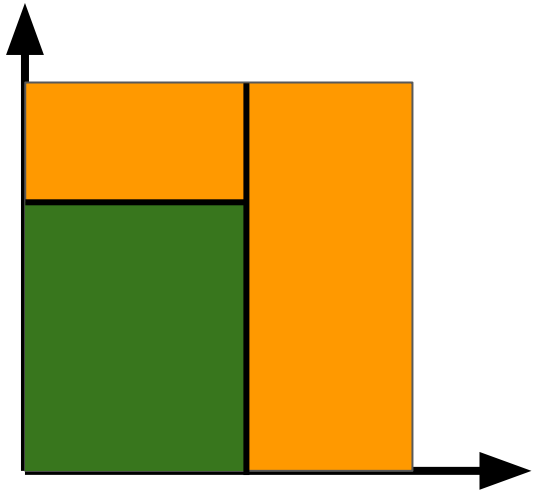


Wider range of hypotheses



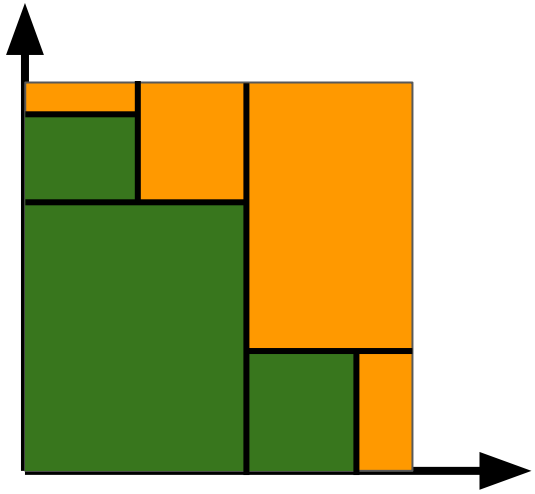
Classification Tree

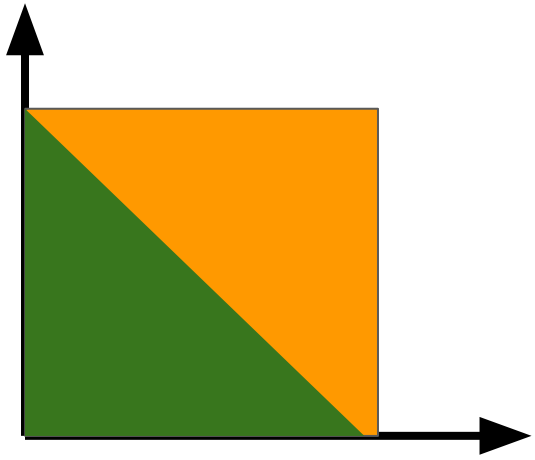
```
if size < 100 then
    if round > 200
        “orange”
    else
        “apple”
else
    “orange”
```



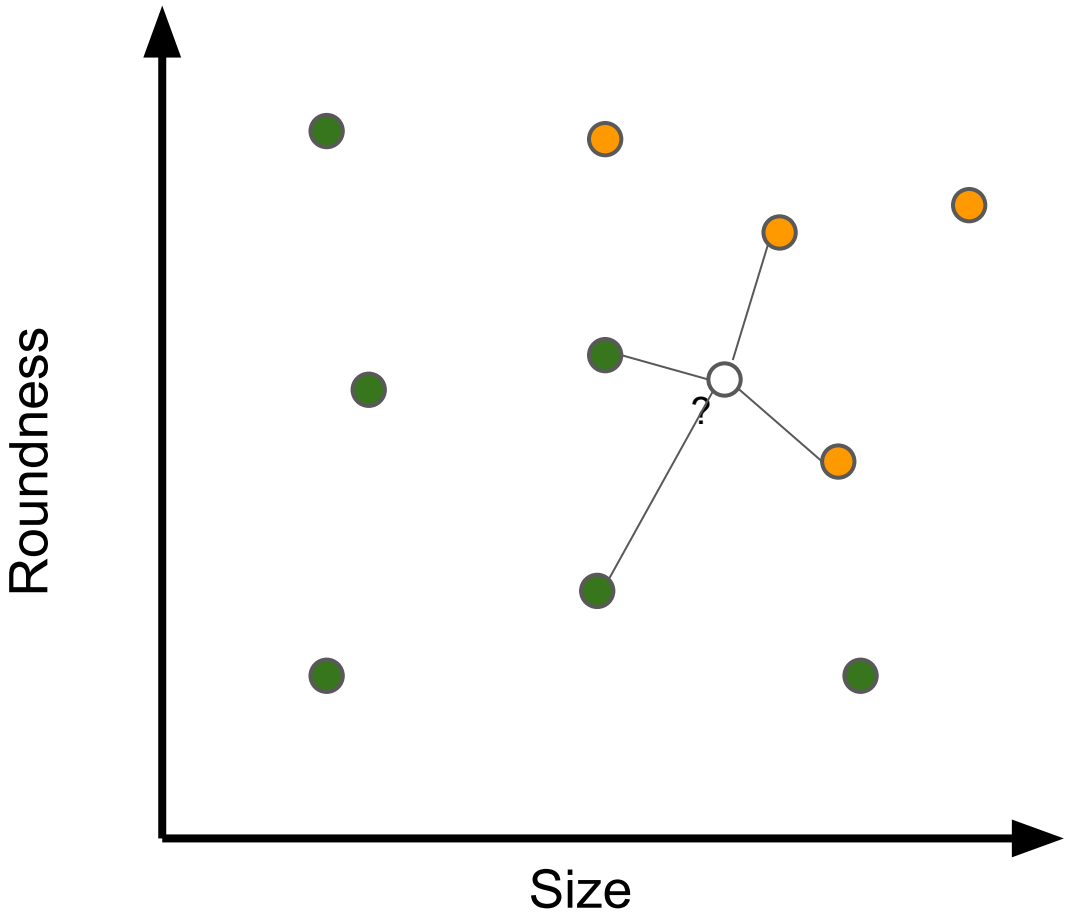
Classification Tree

```
if size < 100 then
    if round > 200
        if size <  50
            if round > 240
                "orange"
            else
                "apple"
        else
            "orange"
    else
        "apple"
else
    if round > 100
        "orange"
    else
        if size < 230
            "apple"
        else
            "orange"
```

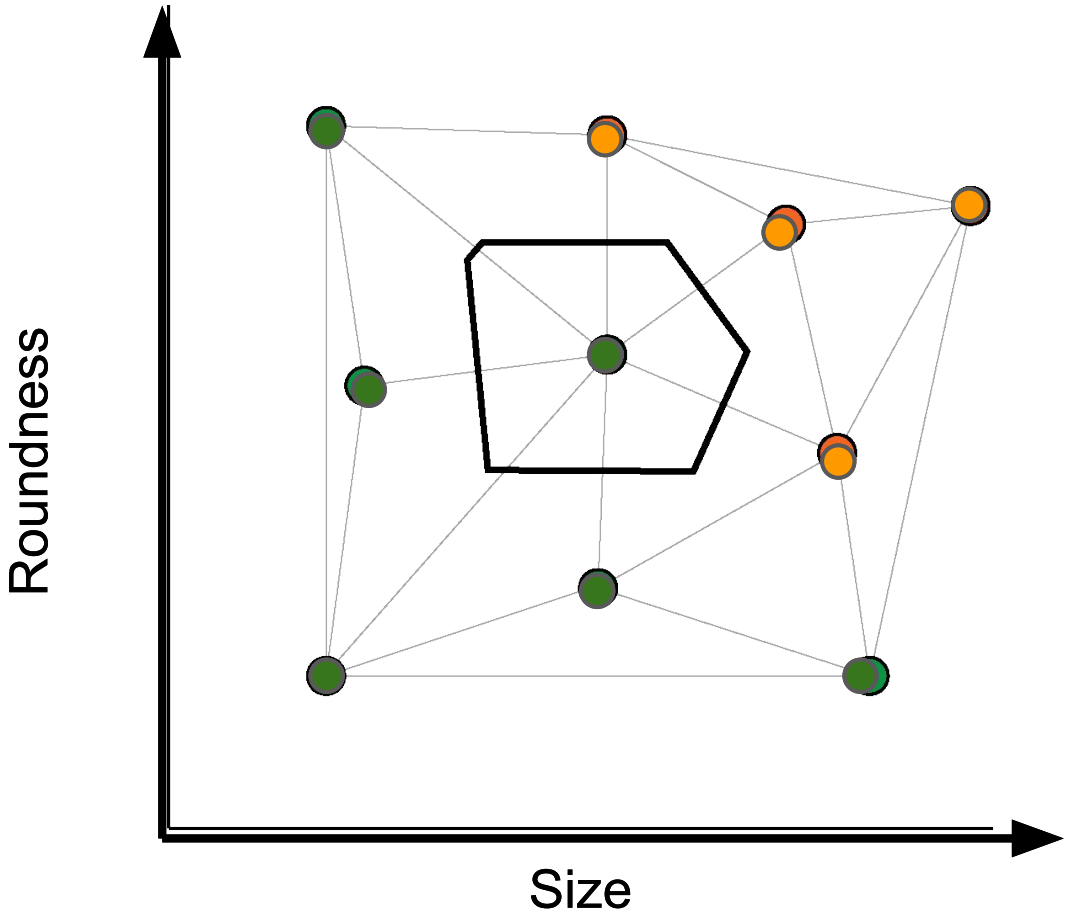




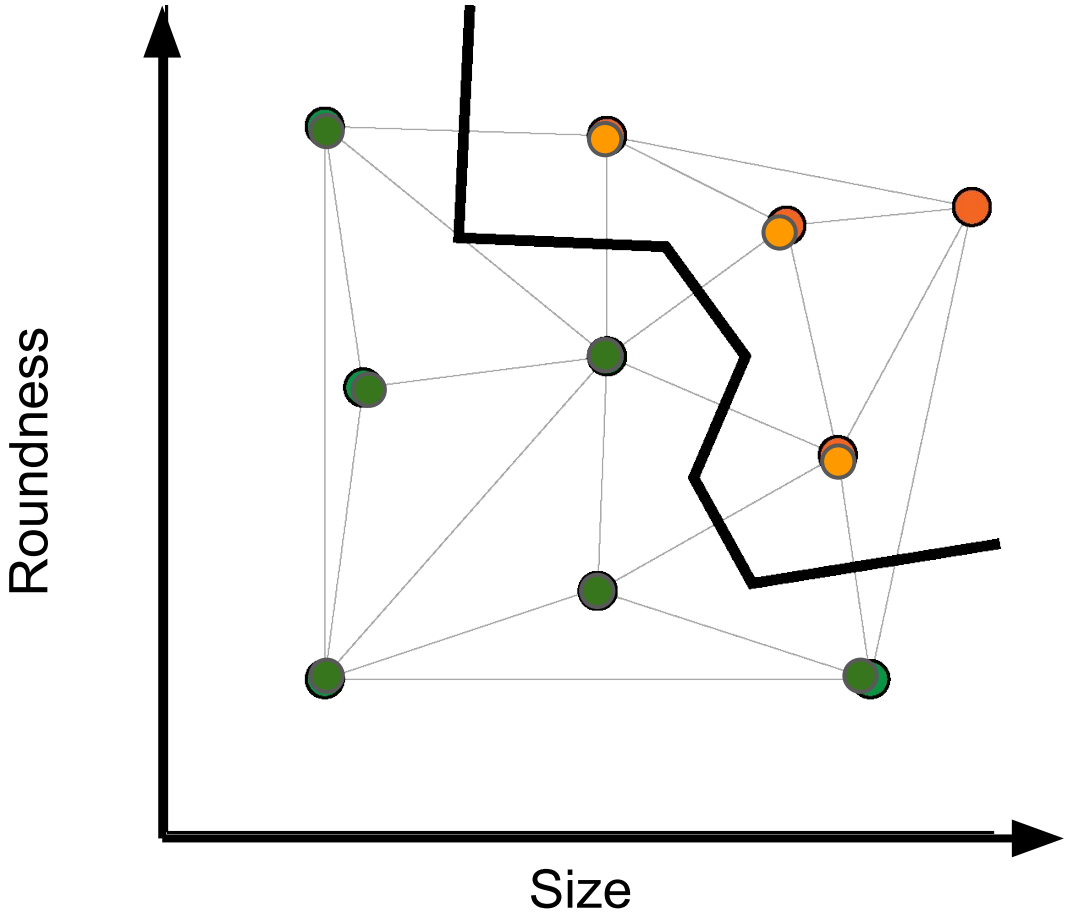
Nearest neighbor



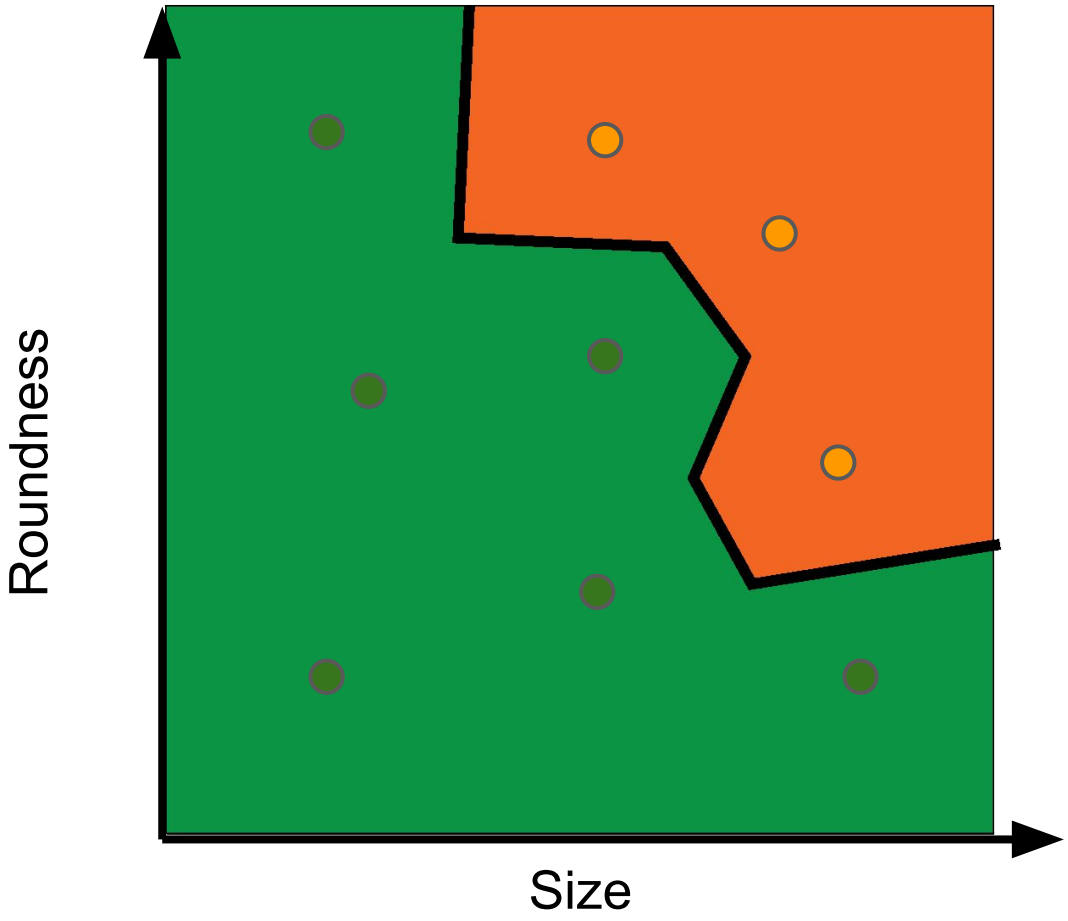
Nearest neighbor



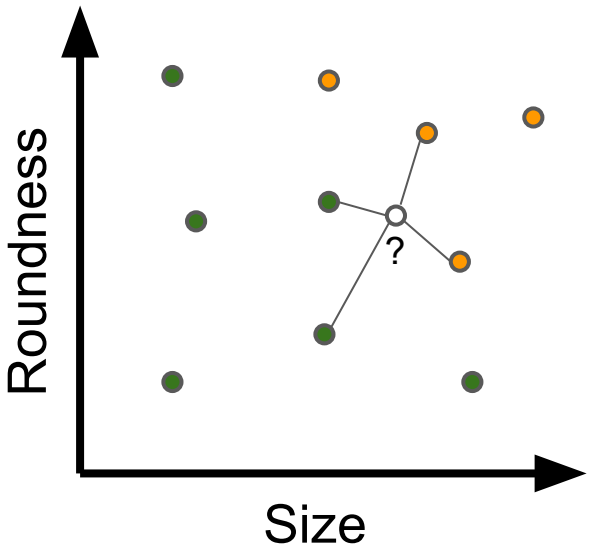
Nearest neighbor

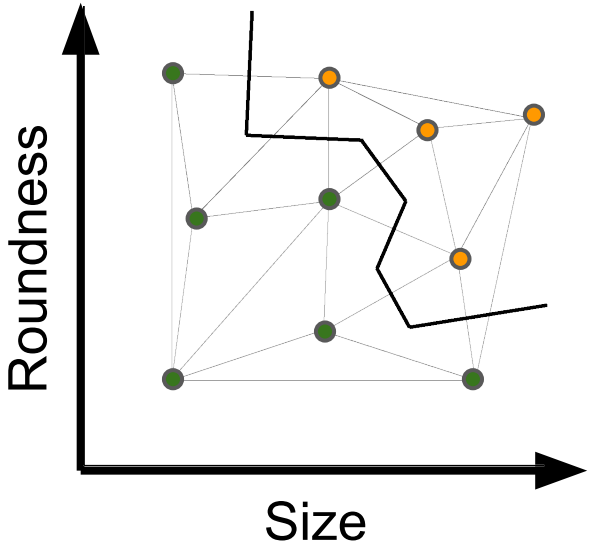


Nearest neighbor



Nearest neighbor





Non-parametric!!

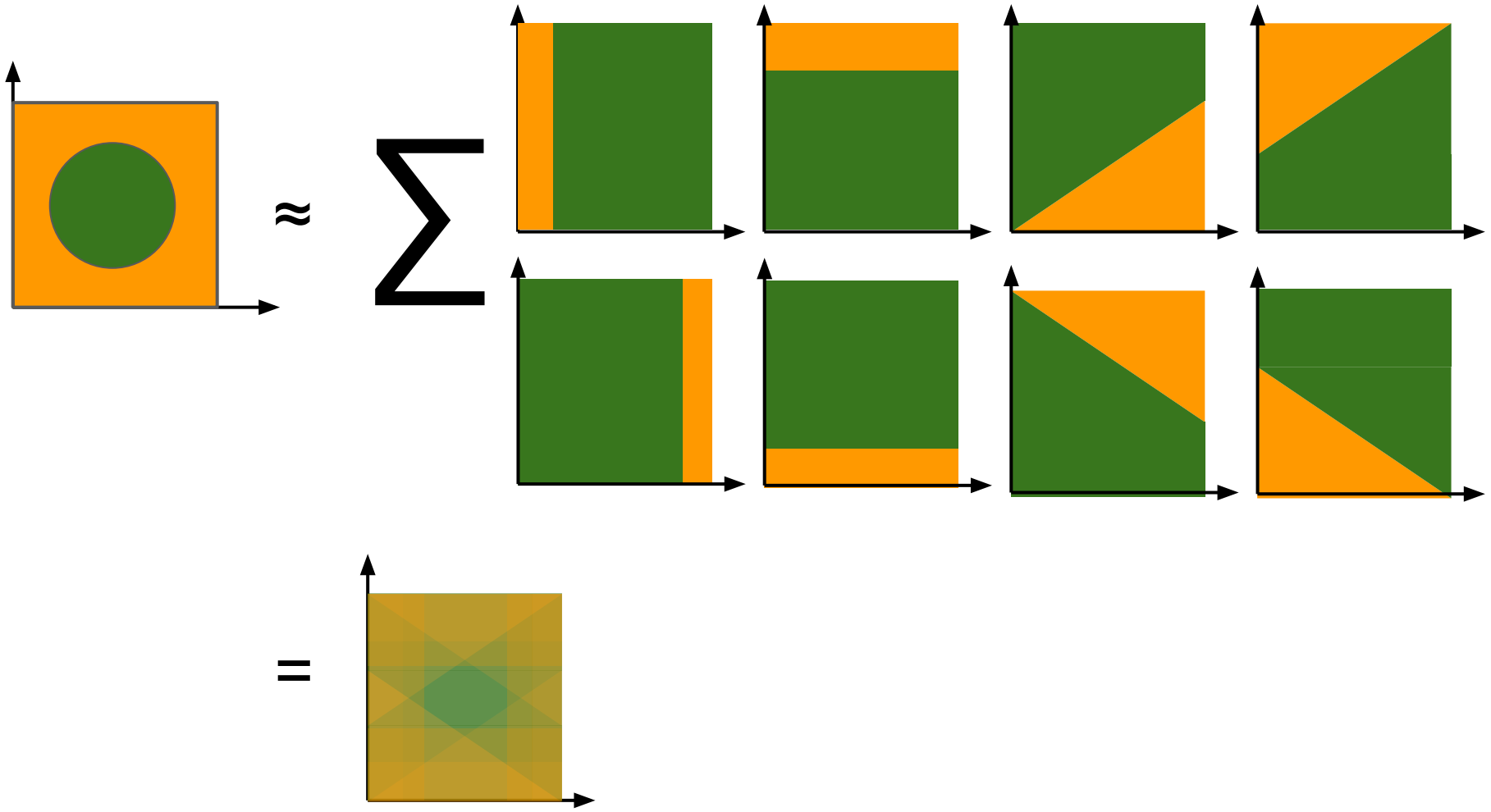
$$\forall i, C_i = C_j | j = \arg\min_k D_{i,k}$$

```
def NN_classify(i)
   j = find_nearest(i)
   return j.class
```

```
def NN_classify(i)
   j = NaN
   min_dist = Inf
   for k in samples:
      if dist(i,k)<min_dist:
         min_dist = dist(i,k)
         j = k
   return j.class
```

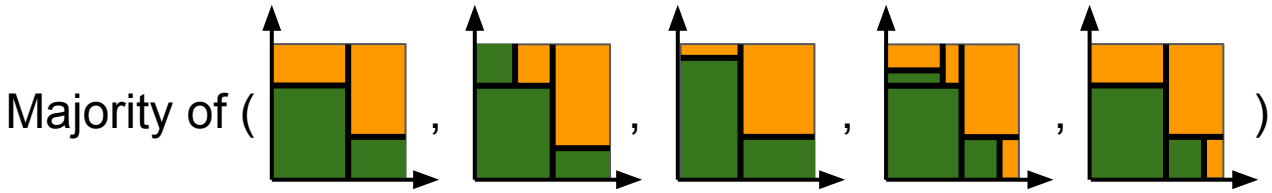
Mixture of experts

- Ensemble
- Wisdom of the crowd
- Non-linear classification landscape



Forest

Ensemble of classification trees

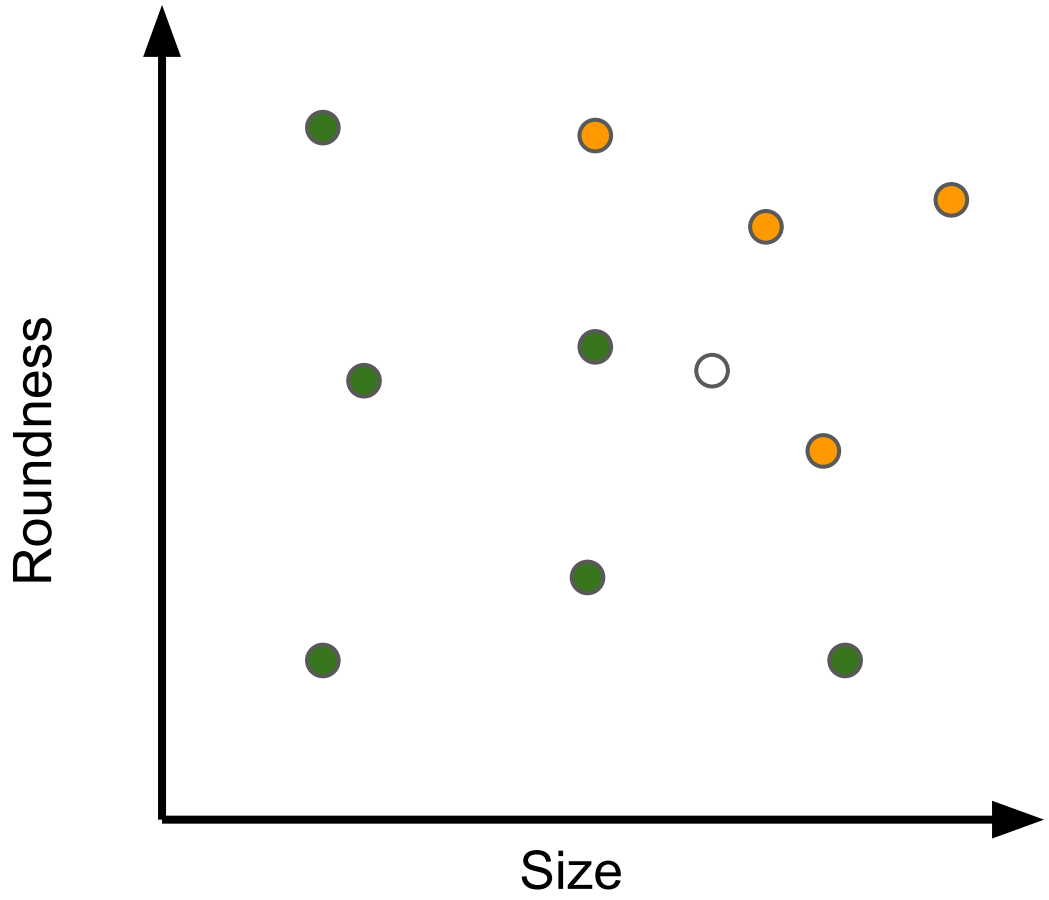


Weighted trees?

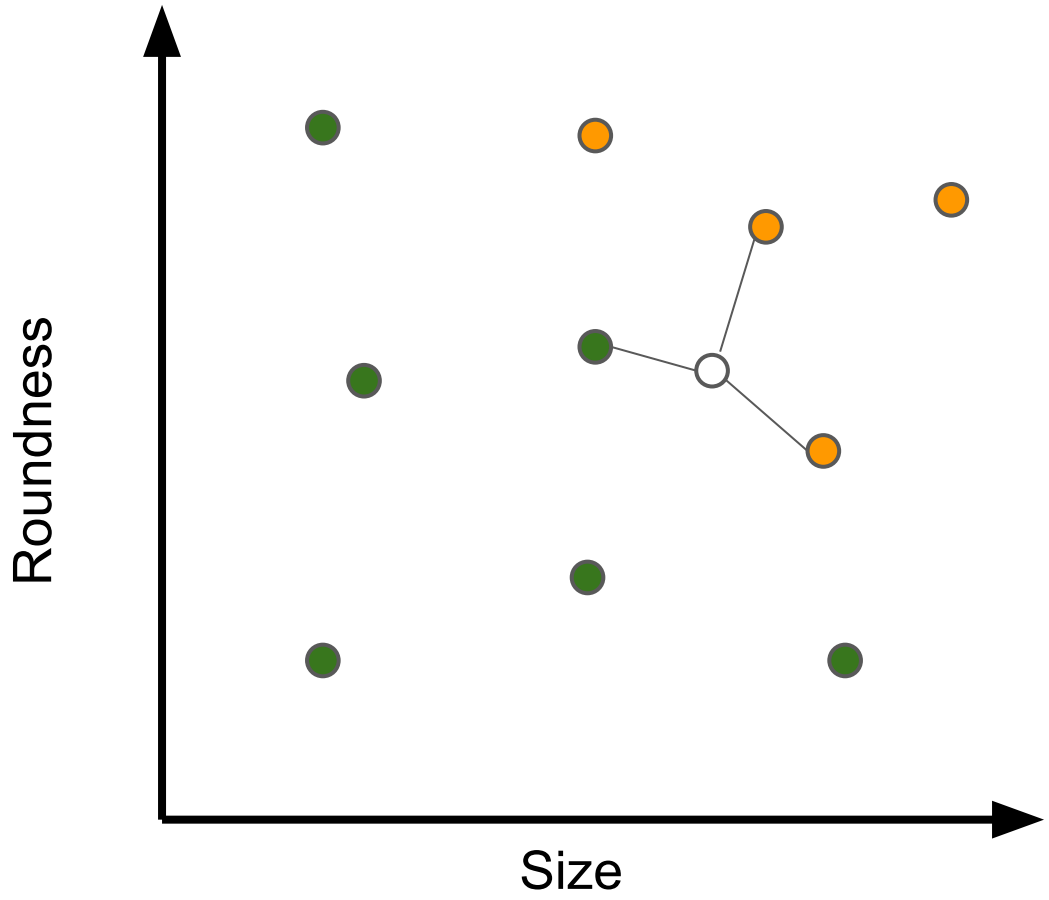
How learned?

Classifiers like: Random forest, Gradient Boosted Trees

K-nearest neighbor

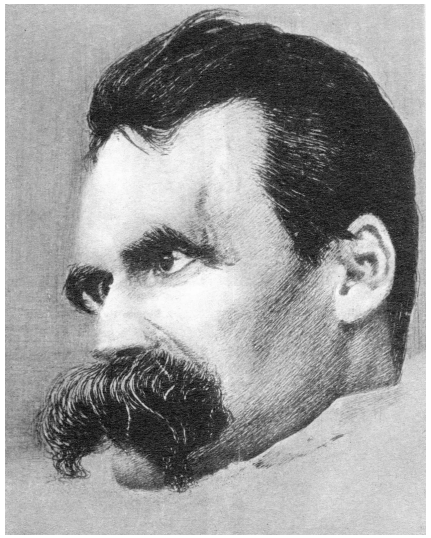


K-nearest neighbor

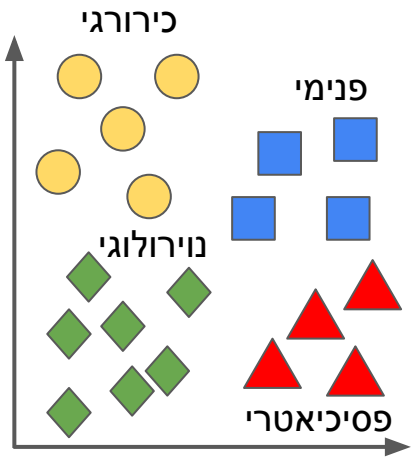


“Beyond Good and Evil…”

Nietzsche

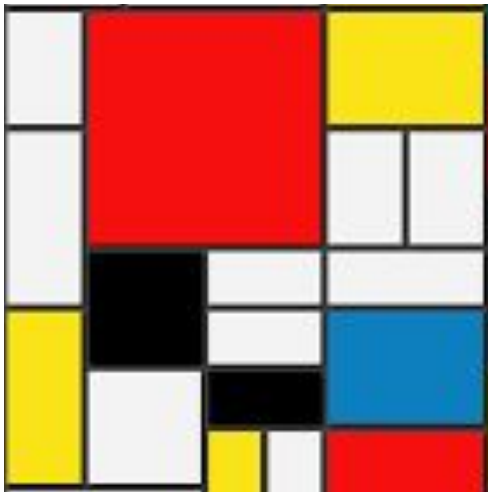


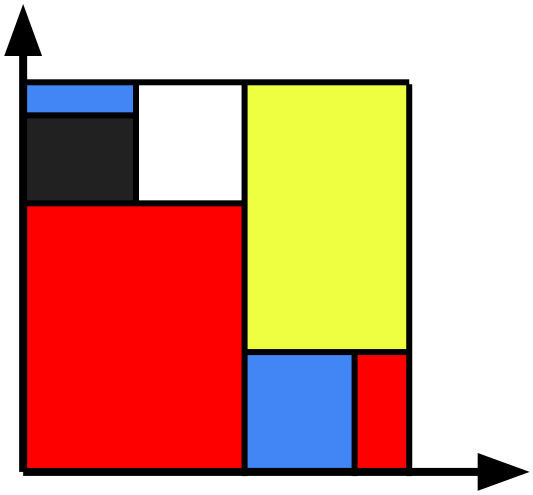
Multiclass classification



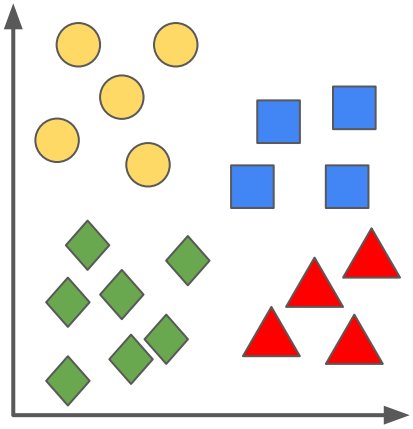
Multiclass classification tree

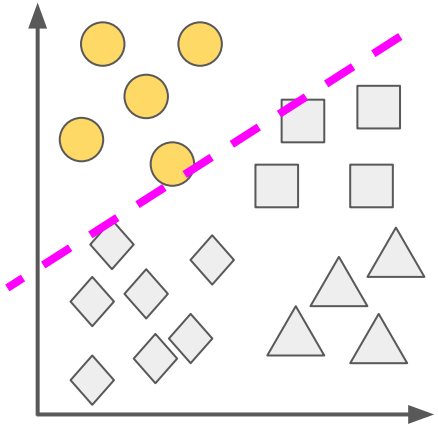
```
if size < 100 then
    if round > 200
        if size < 50
            if round > 240
                "blue"
            else
                "black"
        else
            "white"
    else
        "red"
else
    if round > 100
        "yellow"
    else
        if size < 230
            "blue"
        else
            "red"
```

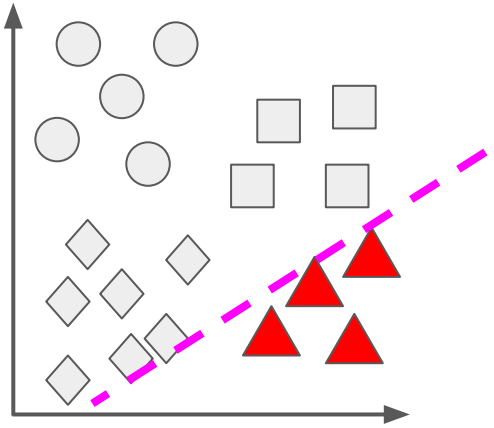


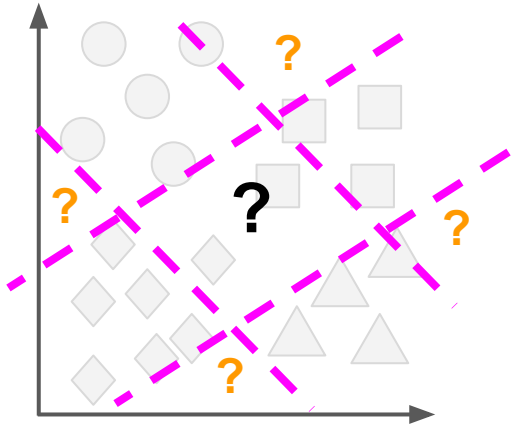


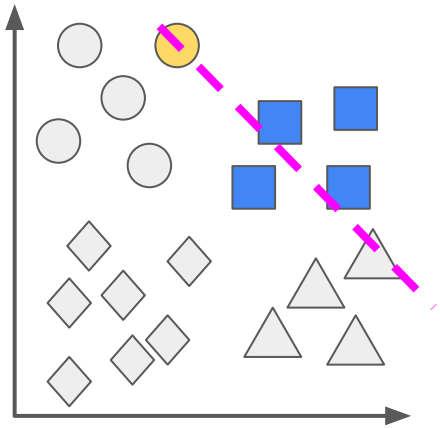
Multiclass classification using binary classifiers
One vs All

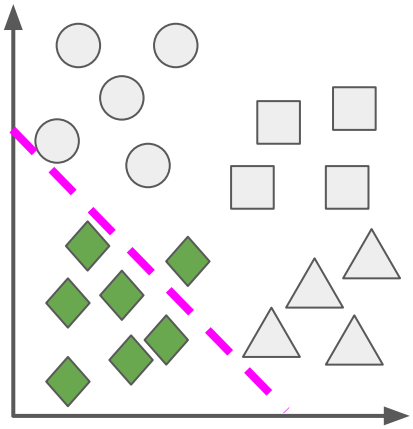




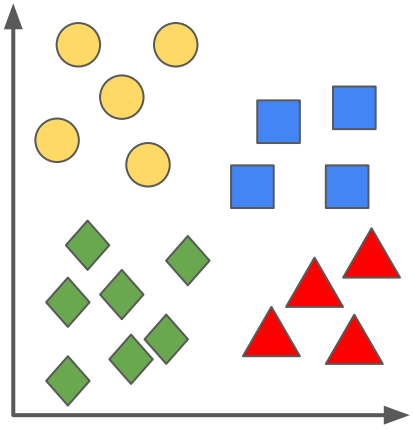


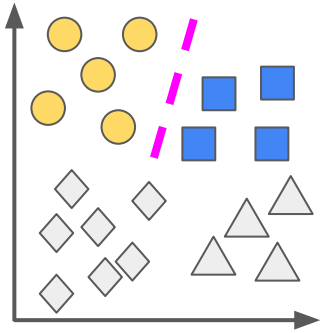


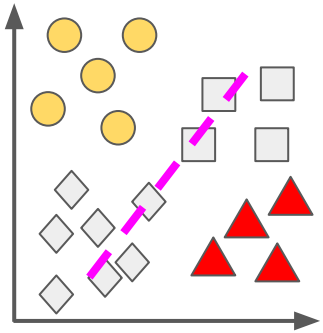


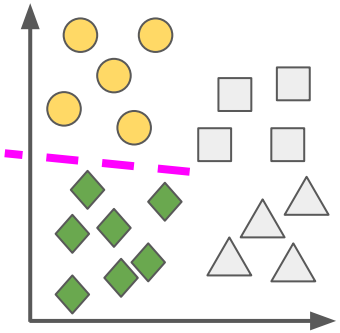


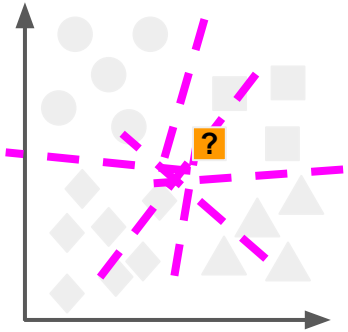
Multiclass classification using binary classifiers
One vs One

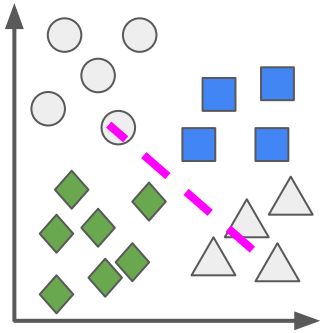


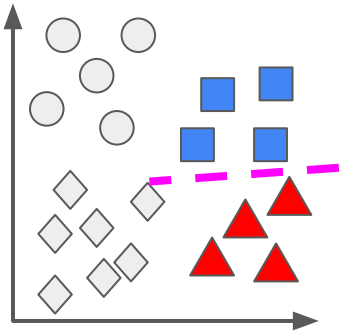


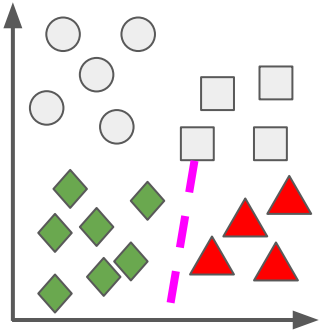


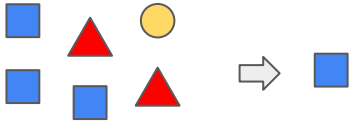












Advanced subjects

- Non-linear classification rules
- Non-linear transformations of the data (generalized linear models)
- Confidence?

Summary

- Classifier
- Different types of classifiers
- Hypothesis class
- Different complexities
- Different types of “rules”
- Capture different aspects of the data

- Ensemble models
- Binary/multiclass classification