
OXMEDICA
Driving Global Education

Applications of Unsupervised Learning

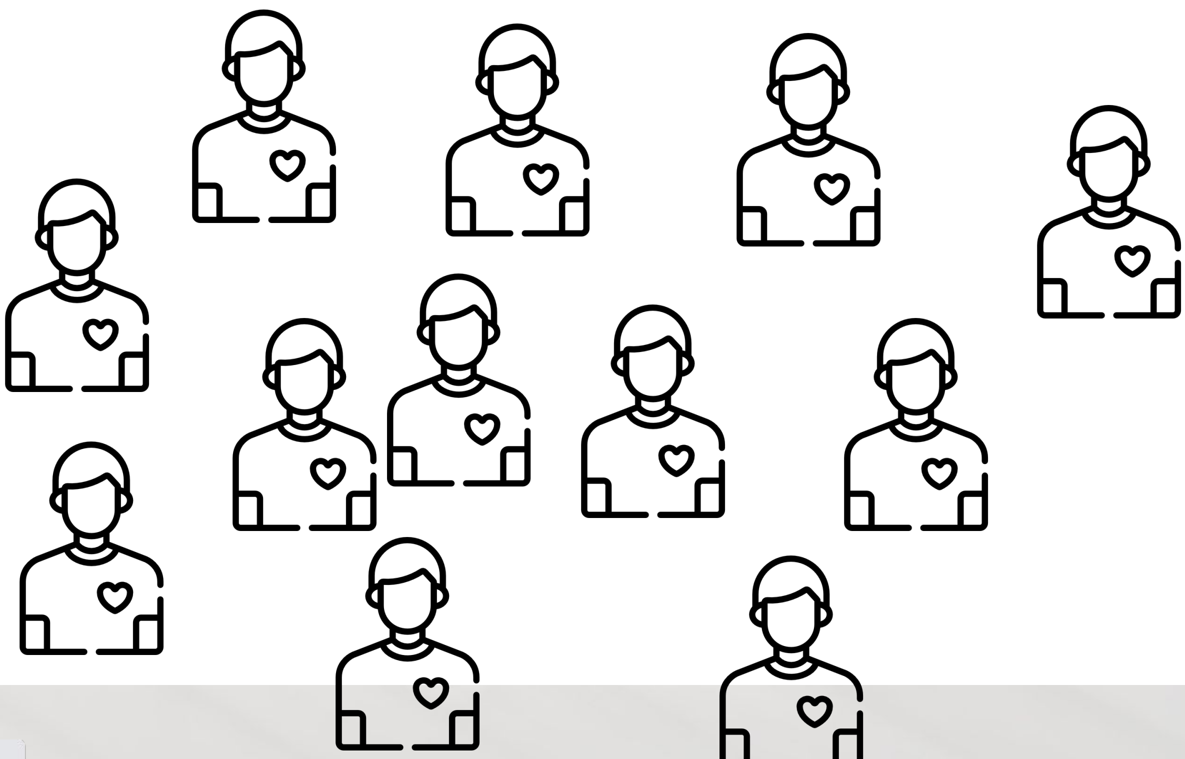
2.07.2024

Example applications of clustering in healthcare and finance

Healthcare example

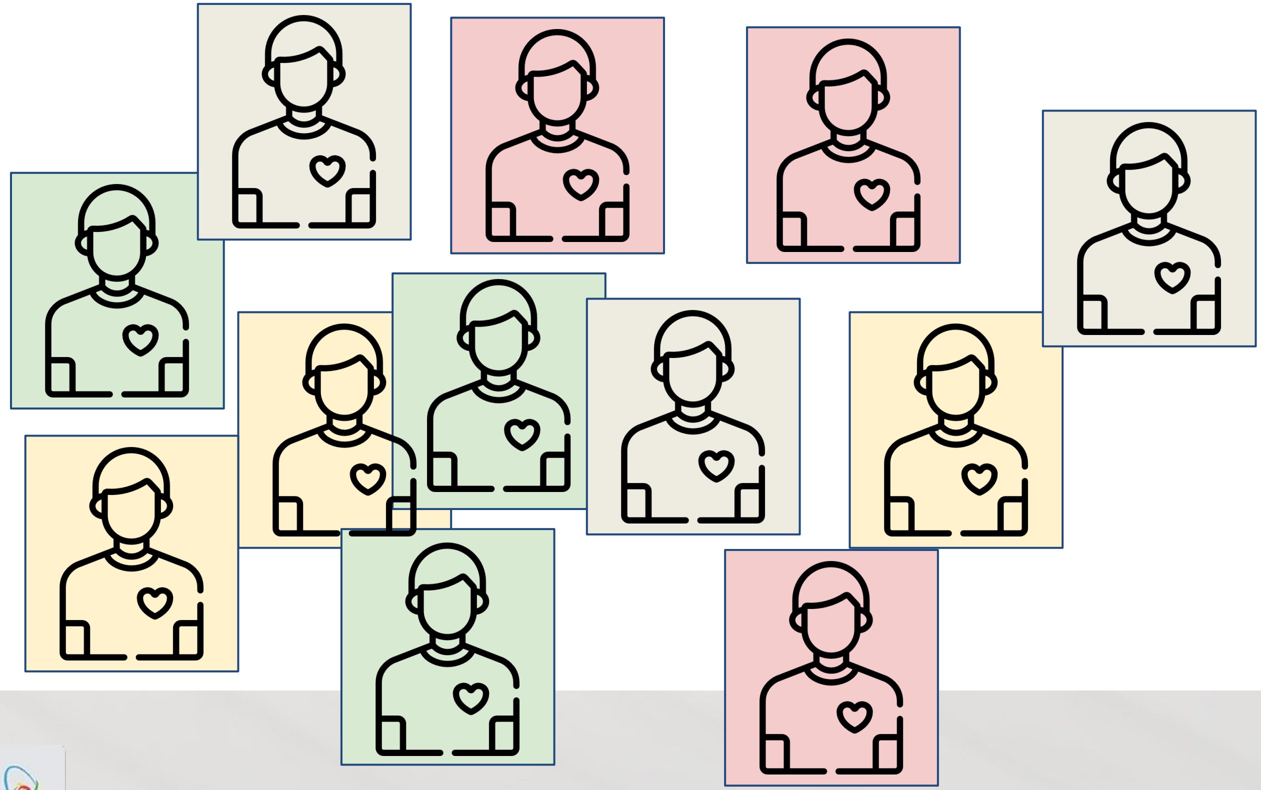
Patient clustering

- Patients may have many different symptoms
- Can use these symptoms to classify them into different risk groups
- This can also be used to automatically detect outliers



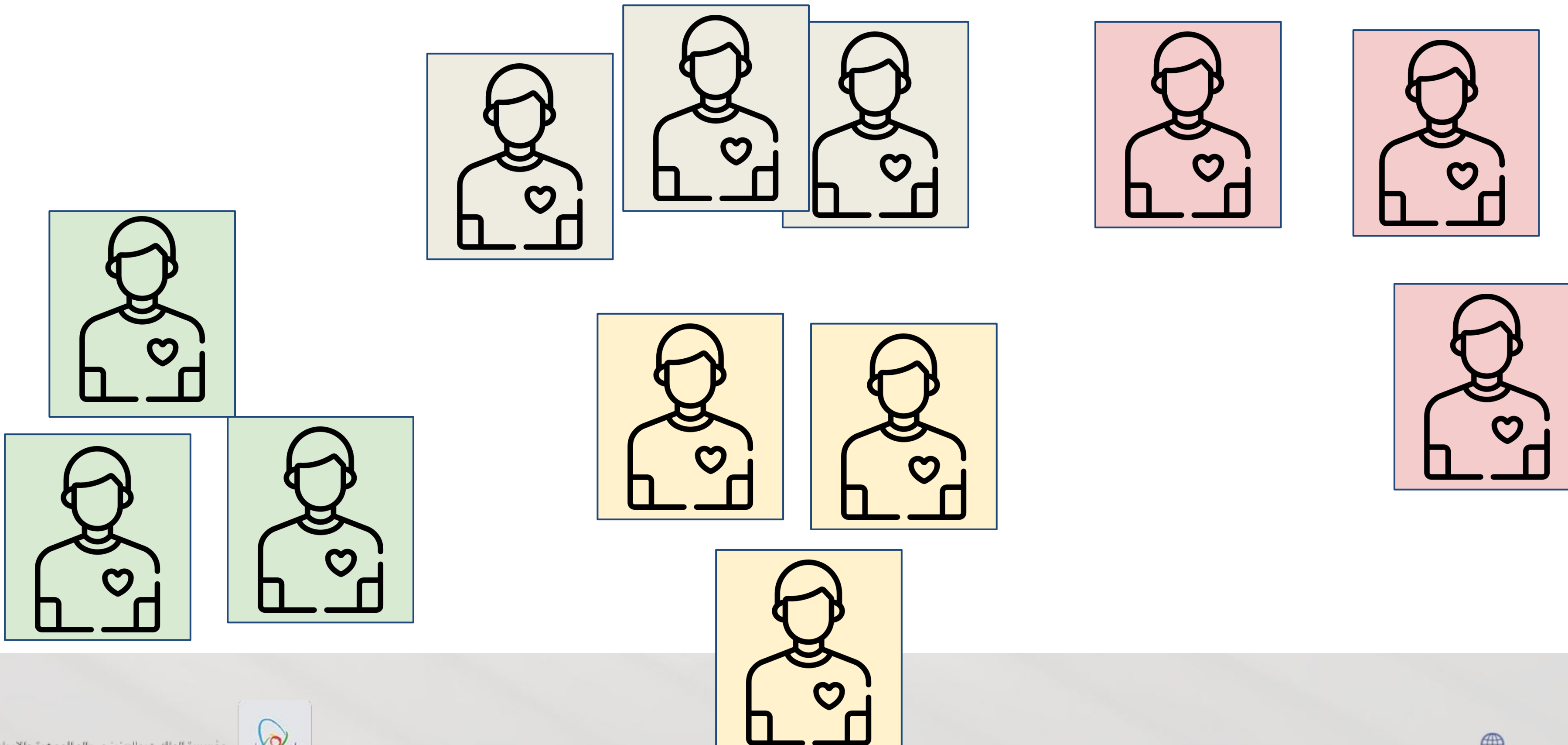
Patient clustering

- Patients may have many different symptoms
- Can use these symptoms to classify them into different risk groups
- This can also be used to automatically detect outliers

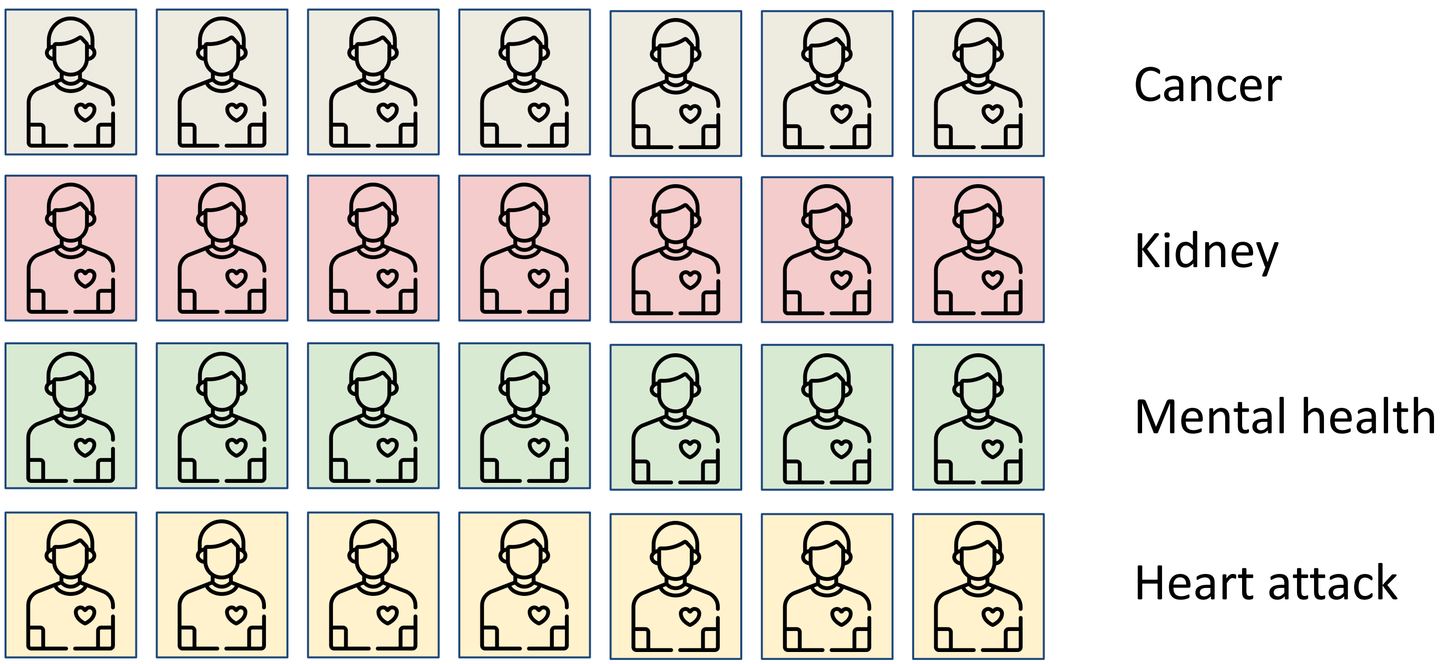


Patient clustering

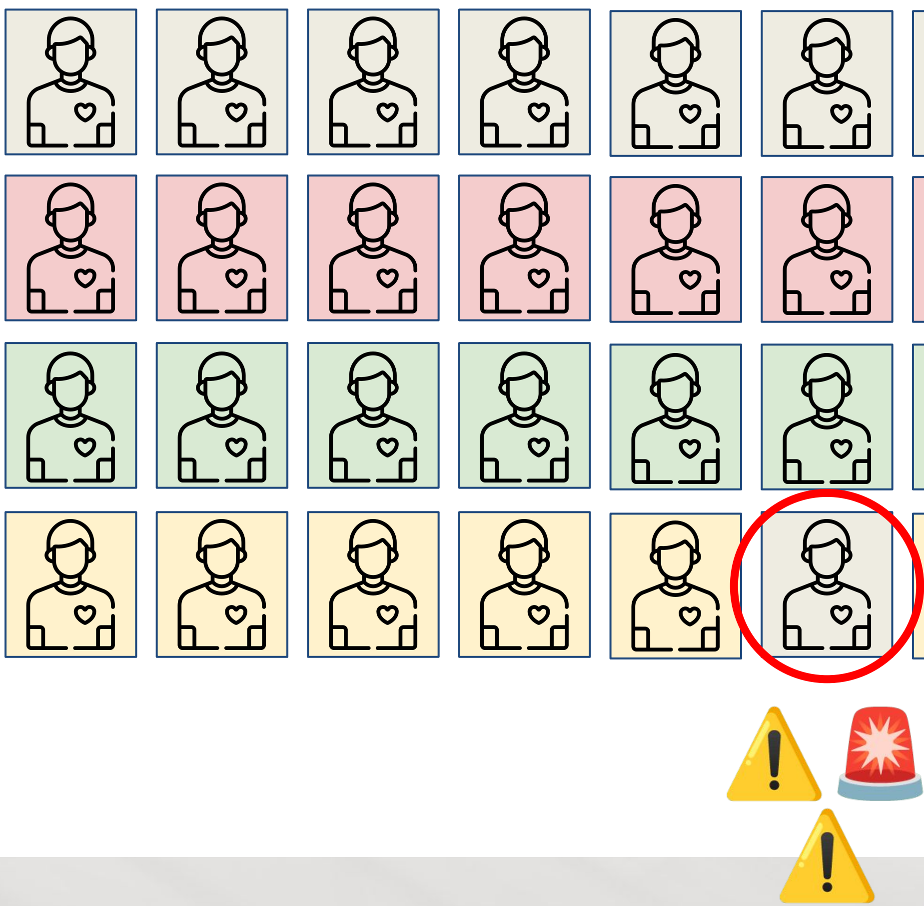
- Use the clusters to help you work out what disease people have
- Useful approach for sub-diseases e.g. helping identify cancer groups



Patient clustering: Can detect outliers



Patient clustering: Can detect outliers



Can help you detect outliers!

Any others?

Finance example

Building a diversified portfolio

- Ideal to build a portfolio of stocks which are have different characteristics to prevent correlated losses.
- Cluster stocks by various metrics e.g. variance, price, risk, industry…etc
- Each group of stocks should be highly correlated within the cluster but have low between-cluster correlation
- Select stocks from different clusters to get a diversified portfolio → lower risk

Building a diversified portfolio

- What kind of risks might this protect the portfolio from?

- What are the limitations of this approach?

Any others?

Clustering in Intensive Care Units

UNSUPERVISED LEARNING APPROACHES FOR IDENTIFYING ICU PATIENT SUBGROUPS: DO RESULTS GENERALISE?

Harry Mayne[1], Guy Parsons[1, 2], and Adam Mahdi[1]

[1]Oxford Internet Institute, University of Oxford
[2]NIHR Academic Clinical Fellow at University of Oxford and Thames Valley Deanery

ABSTRACT

The use of unsupervised learning to identify patient subgroups has emerged as a potentially promising direction to improve the efficiency of Intensive Care Units (ICUs). By identifying subgroups of patients with similar levels of medical resource need, ICUs could be restructured into a collection of smaller subunits, each catering to a specific group. However, it is unclear whether common patient subgroups exist across different ICUs, which would determine whether ICU restructuring could be operationalised in a standardised manner. In this paper, we tested the hypothesis that common ICU patient subgroups exist by examining whether the results from one existing study generalise to a different dataset. We extracted 16 features representing medical resource need and used consensus clustering to derive patient subgroups, replicating the previous study. We found limited similarities between our results and those of the previous study, providing evidence against the hypothesis. Our findings imply that there is significant variation between ICUs; thus, a standardised restructuring approach is unlikely to be appropriate. Instead, potential efficiency gains might be greater when the number and nature of the subunits are tailored to each ICU individually.

https://arxiv.org/pdf/2403.02945

Recent work from my lab at Oxford

Individual Task [10 mins]: Comprehension

- Read the **abstract** of the paper and think about the following questions.

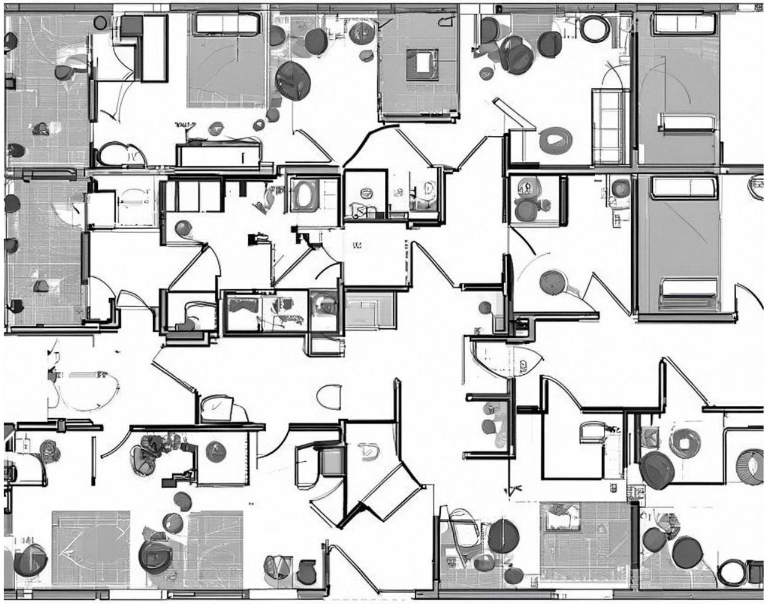
1. What do you think the purpose of an abstract is?
2. What type of unsupervised machine learning does the paper do?
3. What is the specific method used?
4. How many features does it use?
5. What is the aim of the unsupervised machine learning method in this context? I.e. what is the point of the paper?

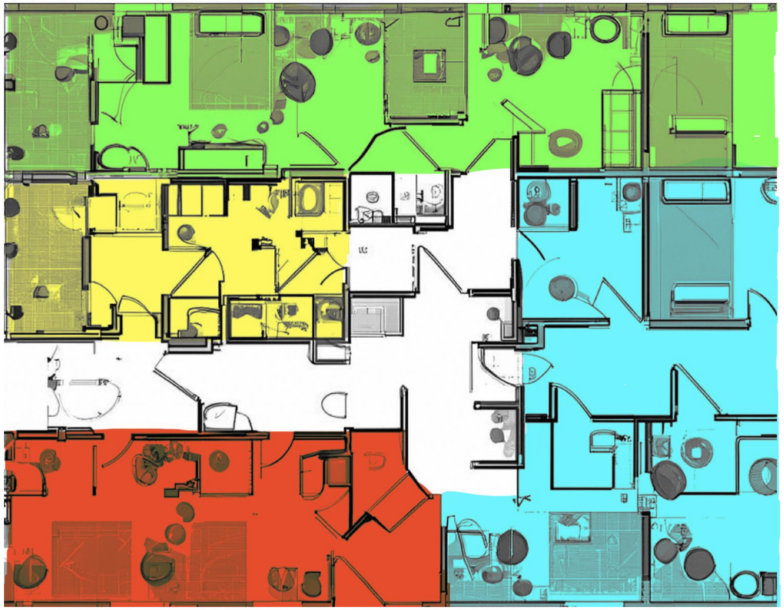
ICU Risks

1. Ageing population
2. Advances in medicine
3. Under-investment

→

- Inefficiency
- Poorer quality of care
- Excessive pressure on clinicians





Group Identification: Results

Cluster 1

48.18%

Relatively healthy

Near perfect survival

Cluster 2

33.68%

Weaker patients

Survive with long-term health problems

Cluster 3

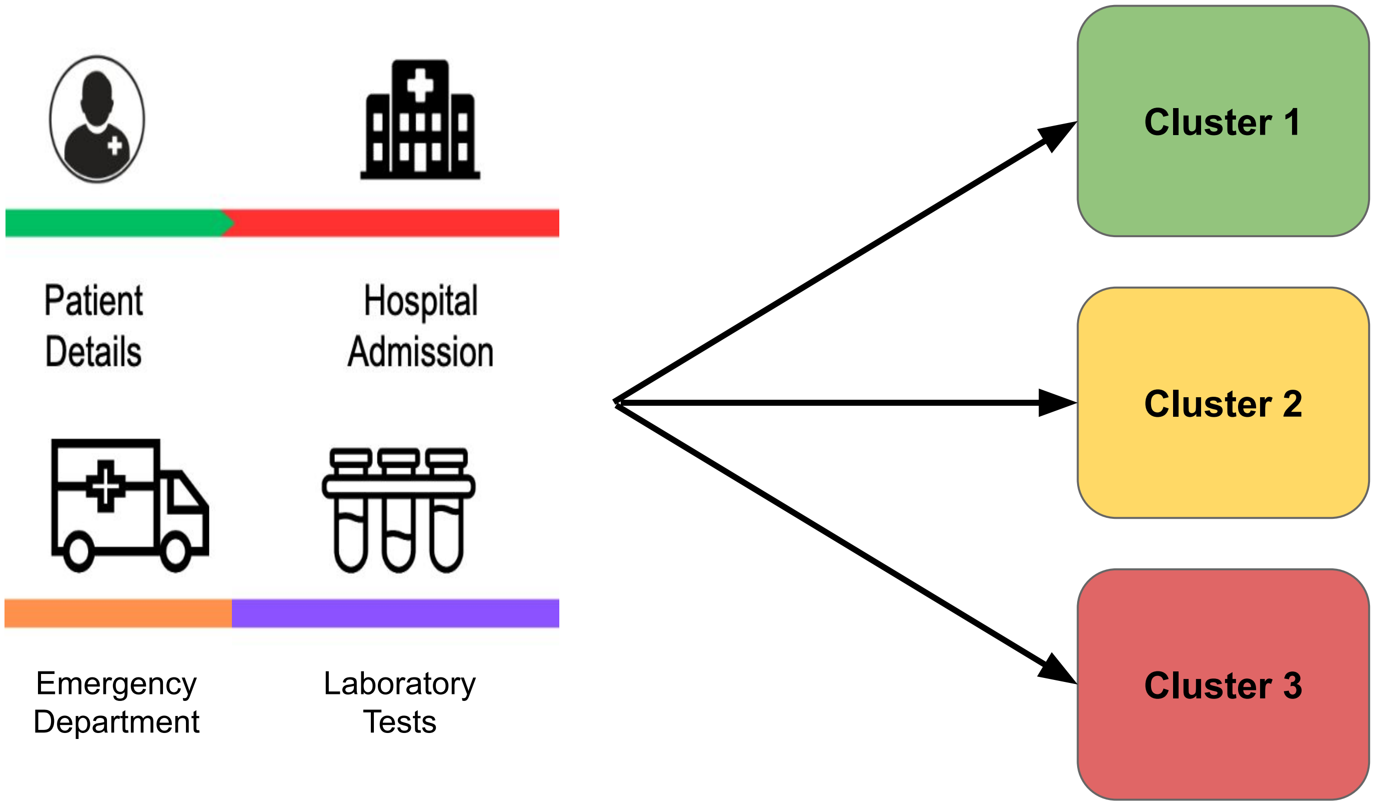
18.14%

Severe patients

76.19% morality

2

Assigning Patients at ICU Admission



Recap questions

1. What are the types of machine learning?
2. What defines unsupervised learning?
3. How can unsupervised learning be used in healthcare?
4. How can unsupervised learning be used in finance?
5. What are the risks of these approaches?

5 Min Break

Supervised Learning

2.07.2024

1 Recap on definitions

2 Notation

3 Regression vs classification problems

4 [EXTRA] Evaluating model performance

5 [EXTRA] A brief introduction to *gradient descent*

Definitions

Definitions Recap

Notation

Supervised Learning Notation

- Covered on the whiteboard
- All notation also uploaded with the **‘Notation’** file on the website

- Predicted values
- Parameters
- Models as functions
- Loss functions

Regression vs Classification Problems

See the whiteboard

Evaluating model performance

See the whiteboard

Loss functions

Discussion

What might be a suitable loss function for a regression problem?

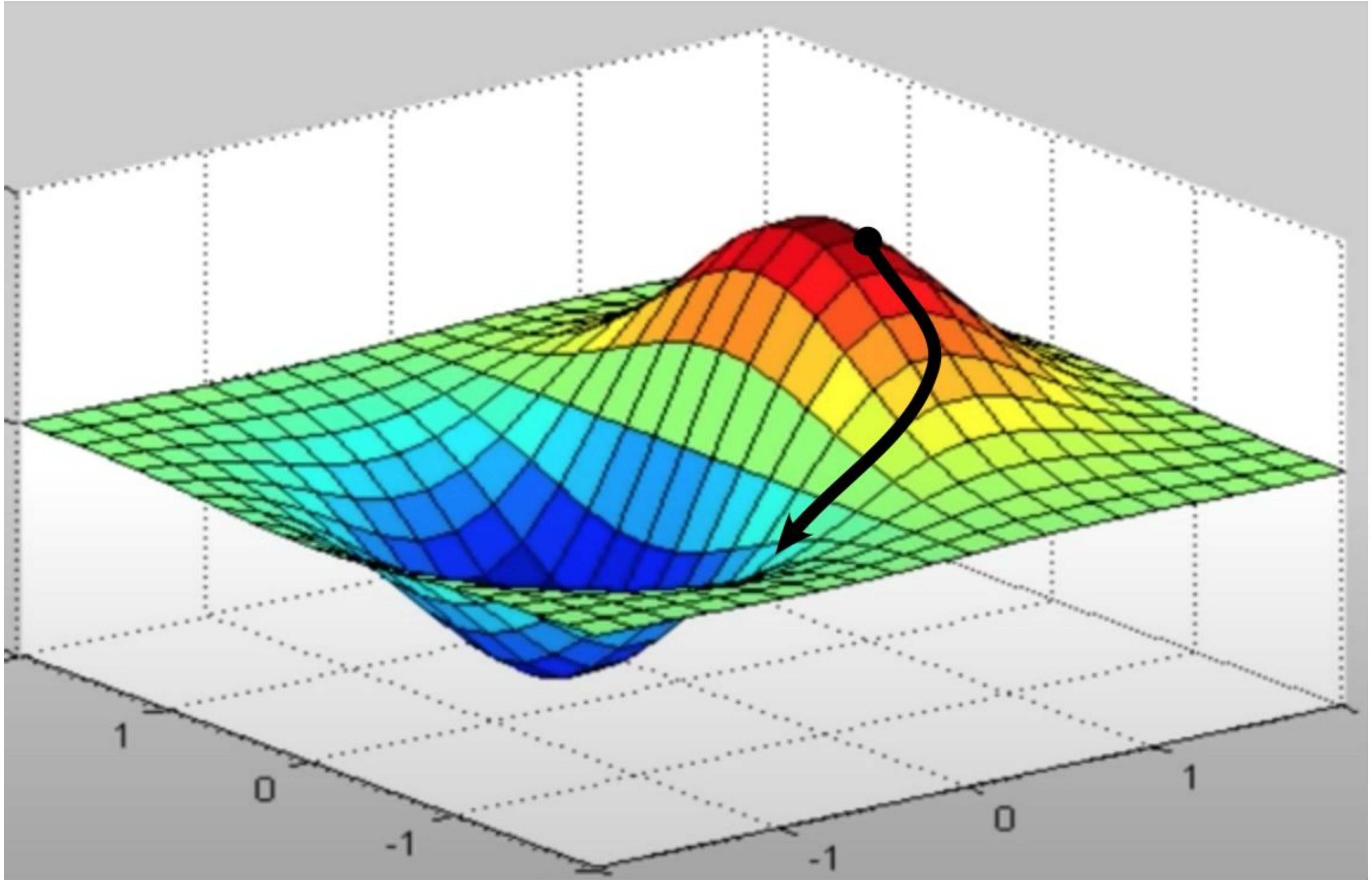
[EXTRA]

Gradient Descent

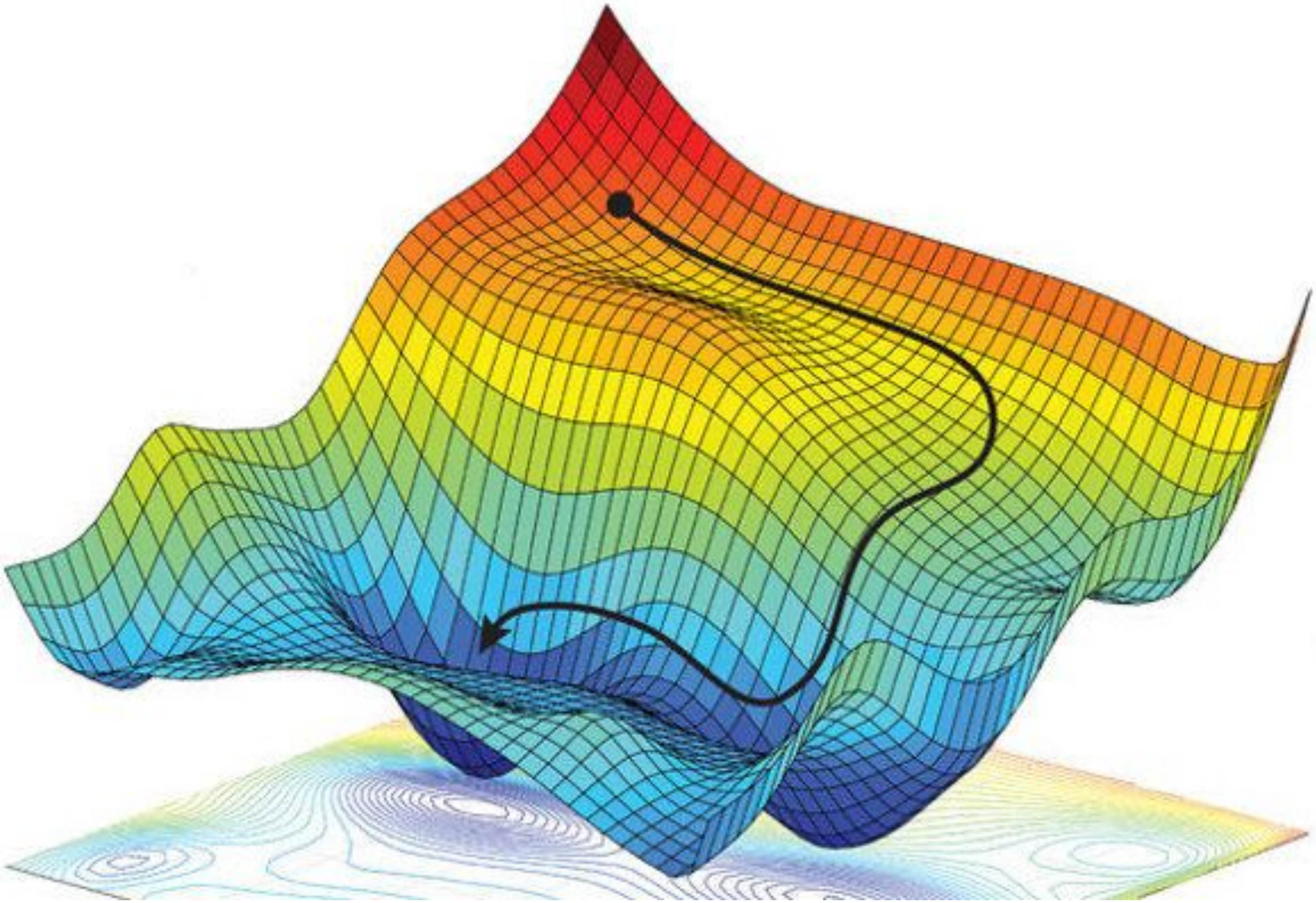
How ‘learning’ actually works…

See the whiteboard

An analogy…

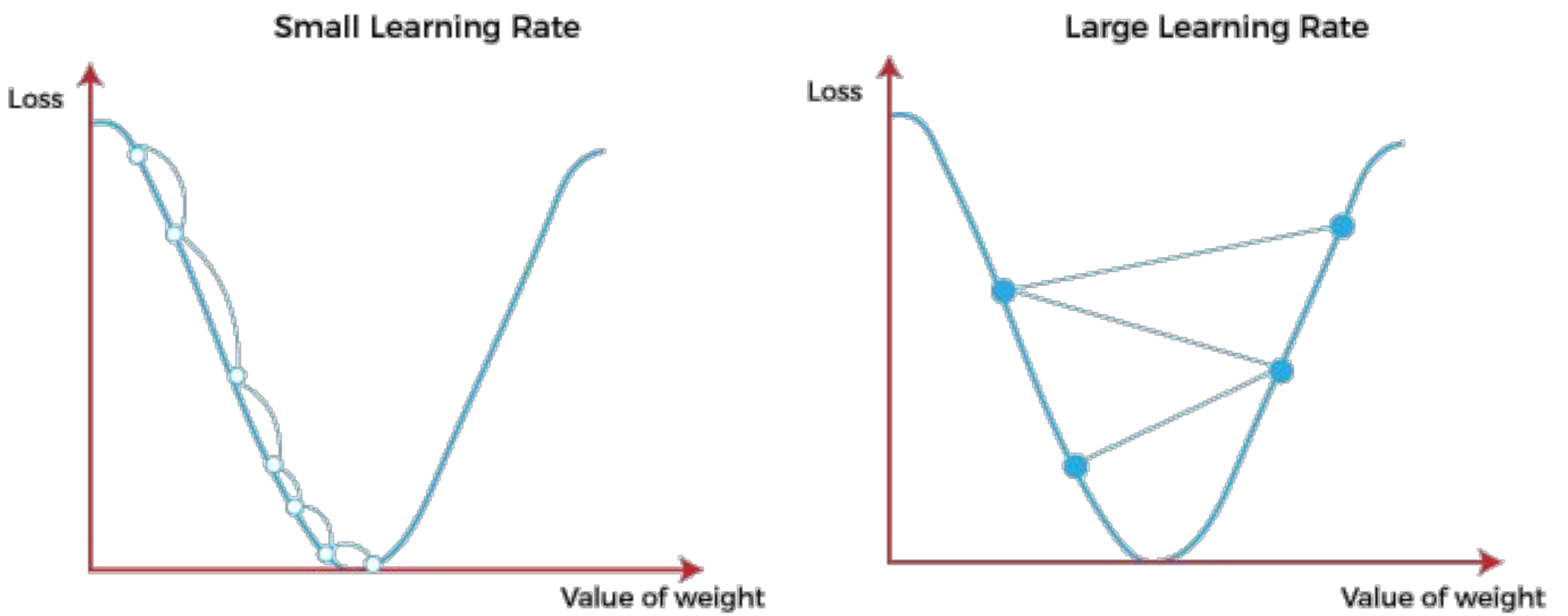


Source



Source

Discussion: Strengths and weaknesses


Small Learning Rate
Loss
Value of weight
Large Learning Rate
Loss
Value of weight

[Source]