Current Topics in AI & ML (Dubai)

Current Topics in Data Science (Dubai)

Week 1: Introduction to the module

(Spring 2024)

Kashif Rajpoot

Outline

1. Learning outcomes
2. Module organisation & delivery
3. Module topics
4. Module distinctiveness
5. Assessment plan
6. Study expectations
7. Quick introduction to AI/ML/DS
8. Research introduction
9. How to read a research paper?
10. Learning outcomes (repeated)
11. MUST DO before week 2 and week 3 lectures
12. Reading group

Learning outcomes

1. Demonstrate an understanding and appreciation of recent advances in AI/ML/DS

2. Make effective oral and written presentations

3. Engage effectively in discussions about recent research

- Let’s discuss each one of them, in latter part of today’s session after the introduction and discussion

Module organisation & delivery

- Edgbaston lectures
  - Pre-recorded lectures (essential to watch for Dubai, except weeks 1 & 2)
  - On-campus lecture recordings (optional for Dubai)
- Dubai lectures: Saturday 10am-2pm
  - Recap of pre-recorded videos (where applicable)
  - Q & A on topics from videos
  - Reading group / discussion continued / tutorial
  - Discussion on selected papers/parts
- Office hours: see Canvas for timing and mode

Module topics and delivery plan

Week	Lecture Date	Time	Lecturer preparing the material	Topic/Title	Watch pre-lecture videos	Delivery in Dubai
1	20.Jan.24	10:00 AM	Kashif Rajpoot & Hamid Dehghani	Introduction and fundamentals	Yes	Kashif Rajpoot
2	03.Feb.24	10:00 AM	Kashif Rajpoot	Trustworthy AI / Responsible AI / Ethical AI	No	Kashif Rajpoot
3	04.Feb.24	10:00 AM	Alex Krull	Denoising in Scientific Imaging	Yes	Kashif Rajpoot
4	10.Feb.24	10:00 AM	Qingjie Meng	Cross-domain Image Analysis	Yes	Kashif Rajpoot
5	17.Feb.24	10:00 AM	Peter Tino	Learning from Time Series Data	Yes	Kashif Rajpoot
6	24.Feb.24	10:00 AM		Mid-Semester Assignment		Kashif Rajpoot
7	02.Mar.24	10:00 AM	Aleš Leonardis	Computational Photography	Yes	Kashif Rajpoot
8	09.Mar.24	10:00 AM	Jinming Duan	Make Machines Learn Better with Prior Knowledge	Yes	Kashif Rajpoot
9	16.Mar.24	10:00 AM	Kashif Rajpoot	Computational Pathology	Yes	Kashif Rajpoot
10	23.Mar.24	10:00 AM	TBA	Guest Lecture in NLP	?	Kashif Rajpoot
	30.Mar.24			Easter Break		
	06.Apr.24			Easter Break		
	13.Apr.24			Easter Break		
	20.Apr.24			Easter Break		
11	27.Apr.24			End-Semester Assignment and Peer-Review		

Module topics: week 2

- Dubai specific
- Potential risks and harms of AI
- AI Ethics
- Trustworthy AI
- Responsible AI
- Ethical AI

Module topics: week 2

- Edgbaston specific: optional for Dubai campus

Hyung Jin Chang

- Vision-based human body pose, human hand pose, eye gaze, and 6D object pose estimation methods.

Module topics: week 3

Alexander Krull

- Concept of 'Denoising in Scientific Imaging,
- As applied in Biomedical Sciences through Fluorescence Microscopy to account for Noise

Module topics: week 4

Qingjie Meng

- Domain shift in machine learning
  - Cross-domain image analysis.
  - Implications for medical imaging analysis.
  - Cross-domain ultrasound classification and cross-domain MRI segmentation.

Module topics: week 5

Peter Tino

- Issues that emerged in Machine Learning when the data exhibits temporal dependencies.
- When data cannot be considered in isolation and the order in which the data is presented matters, e.g. stock price prediction

Module topics: week 7

Ales Leaonardis

- Computational Photography
  - Methods for computer vision, object and scene recognition and categorization, and object tracking.
  - Using computing techniques such as artificial intelligence, machine learning to capture images.

Module topics: week 8

Jinming Duan

- Leveraging prior knowledge such that machine can learn better.
- Will cover some recent advances of the usage of machine (deep) learning in medical imaging.

Module topics: week 9

Kashif Rajpoot

- Computational Pathology
  - AI and machine learning algorithms for the study of histological and multi-omic markers of cancer biology,
  - early detection of cancer and stratification of cancer patients in terms of recurrence, progression and response to therapy

Module topics: week 10

- Natural Language Processing

Module distinctiveness

- What's different about this module?
  - Not just one-way lectures on established book concepts
  - Get exposure to AI/ML/DS research and applications (with Birmingham perspective)
  - Develop ability to look and study about research in AI/ML/DS
  - In-class discussions and reading groups

Assessment plan (tentative)

- Formative assessment – no marks, just for practice!
  - Weekly quiz (Canvas and/or in-lecture) starting from Week 2
- Summative assessment – marks count!
  - Mid-Semester Assignment, Week 6 (40%)
    - Submit 2-page review of an assigned paper
    - Release: 15th Feb; Due: 26th Feb.
  - End-Semester Assignment, Week 11 (60%)
    - Submit 5-minutes recorded video presentation containing ‘critical review’ of an assigned topic / area (40%)
    - Peer assessment of 5 presentations (10%) (to be done during Week 11)
    - Peer-reviewed grade (10%)
    - Release: 18th Mar; Due: 22nd Apr.

Study expectations (1 of 3)

- Who is studying this module and with what AI/ML/DS background?
  - MSc AI & CS (conversion MSc),
    studying AI & ML in parallel or studied it last year
  - MSc DS (conversion MSc for numerical background),
    studied Algo for DS before
  - MSc AI & ML (specialist MSc), studied ML and/or NC before
- What can you expect/learn from this module?
  - Get exposure to AI/ML/DS research and applications
  - Develop ability to look and study about research in AI/ML/DS
  - It is a module which can become very personalised based on your knowledge background and experience
    - Set yourself a goal to learn about the breadth and depth of research

Study expectations (2 of 3)

- What's the format of the module?
  - Pre-lecture: watch pre-recorded videos at home (where applicable as pointed out by module lecturer)
  - In-lecture: attend on-campus lectures, and participate in reading group and discussions
  - Post-lecture: read/watch recommended reads/videos
- ~10-12 hours of study time / week
  - ~2 hours of watching pre-recorded videos, before on-campus lectures
  - ~4 hours of on-campus lecture
  - ~4-6 hours of reading / watching
- Bring a device (e.g., laptop or tablet)
  - You will need to read (digitally or take print and then read)
- Join on time
- Ask questions / participate i.e. make it an interactive module

Study expectations (3 of 3)

- What Maths background is needed?
  - Several topics and research papers will be using strong Maths
  - What to do if you're not familiar/fresh with Maths?
    - Don't panic! Maths is inevitable in AI/ML/DS research.
    - Study Maths bootcamp (self-study on Canvas) and introductory material in initial weeks of Current Topics
    - Filter maths (where needed, for conversion MSc)
  - Feel free to ask for advice from module lecturer
- Develop ability to distinguish hype from reality by understanding limitations of a research study
- Join Dubai team UoBD PG CS 2023 (on Microsoft Teams) and then Current Topics channel: to be used for questions, updates, announcements, etc

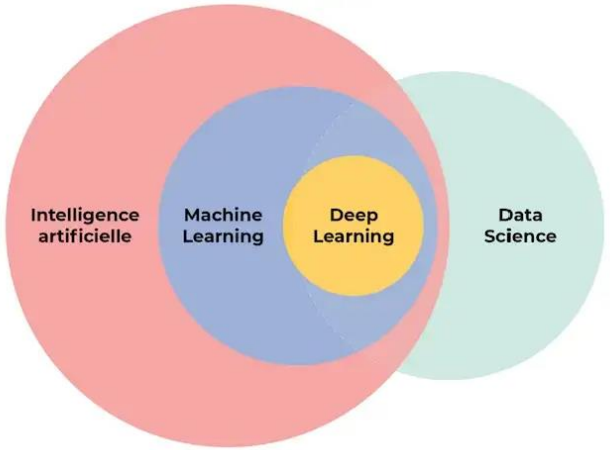
Quick introduction to AI/ML/DS (1 of 4)

- ChatGPT generated definitions of AI/ML/DS
  - Artificial Intelligence is the simulation of human intelligence in machines designed to think and act like humans. It involves using data and machine learning algorithms to enable computers to perform tasks requiring human-like intelligence, such as decision-making, language understanding, and pattern recognition.
  - Machine Learning is a subfield of AI that allows computer systems to improve their performance on a task by learning from data, without explicit programming. It involves building models that can make predictions, classify data, or identify patterns based on statistical analysis of data.
  - Data Science is the field of using scientific methods, algorithms, and systems to extract insights and knowledge from data. It involves analysing, cleaning, and modelling data to make predictions and inform decisions.

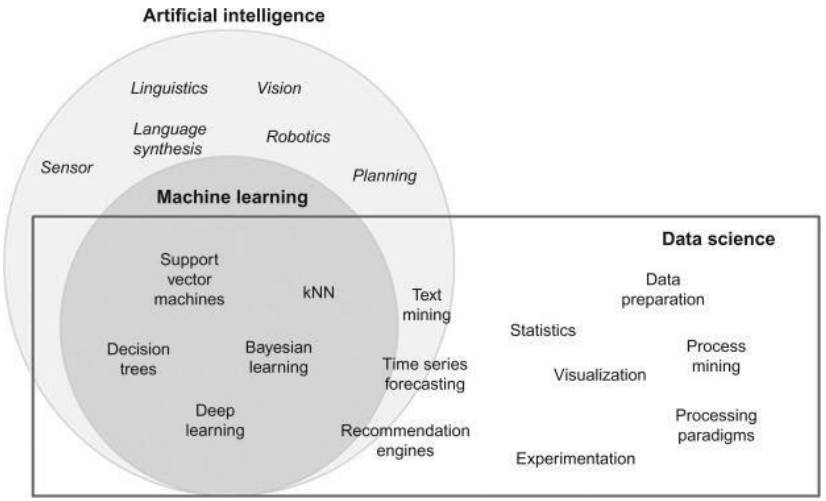
Quick introduction to AI/ML/DS (2 of 4)

- ChatGPT generated distinction between AI/ML/DS
  - AI, ML, and DS are interrelated but distinct fields:
    - AI is the broader concept of machines being able to perform tasks that normally require human intelligence.
    - ML is a subset of AI that focuses on the ability of machines to learn from data and improve their performance on a task without being explicitly programmed.
    - DS is a field that uses scientific methods, processes, algorithms, and systems to extract insights and knowledge from structured and unstructured data. It involves using data and machine learning to inform decisions and build predictive models.
  - In summary, AI involves building machines that can perform tasks requiring human intelligence, while ML involves teaching these machines to learn from data. DS is the field of using data to extract insights and inform decisions, and it often utilizes ML techniques to build predictive models.

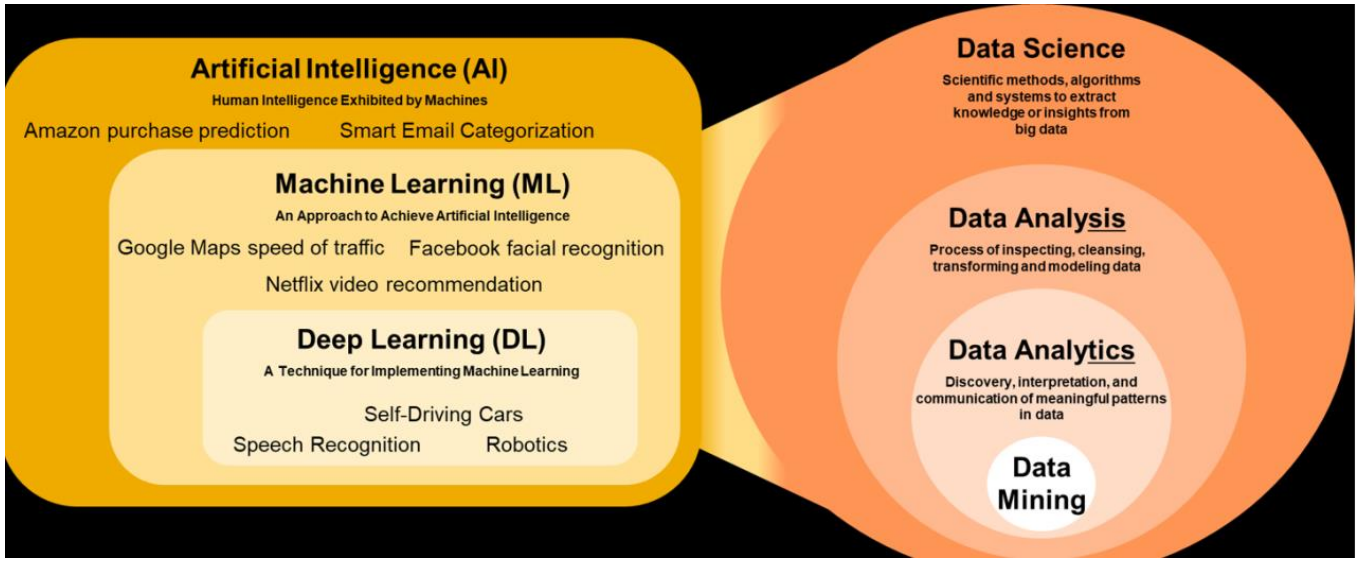
Quick introduction to AI/ML/DS (3 of 4)



https://medium.com/geekculture/machine-learning-data-science-and-artificial-intelligence-a45a2ffe9639



https://www.sciencedirect.com/topics/physics-and-astronomy/artificial-intelligence



https://www.linkedin.com/pulse/how-make-simple-explain-ai-ml-dl-together-data-science-vollmer

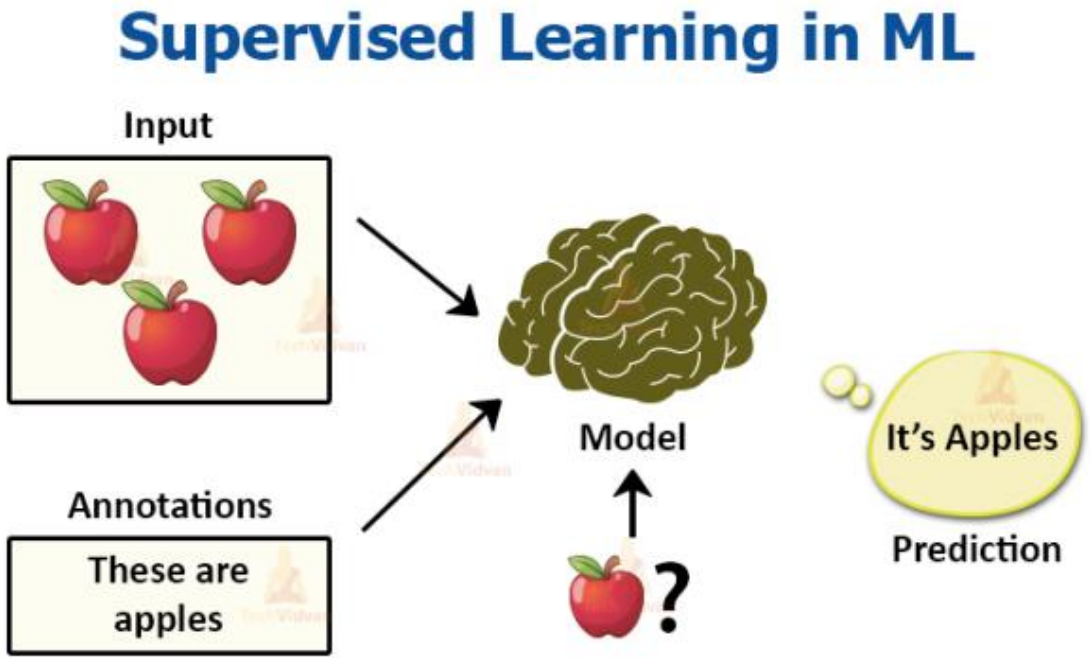
Quick introduction to AI/ML/DS (4 of 4)

- Where to learn fundamentals of AI/ML/DS?
  - Your respective modules
  - https://towardsdatascience.com/
  - https://medium.com/tag/data-science
- Examples:
  - https://medium.com/almabetter/machine-learning-fundamentals-for-beginners-70f409b1197e
  - https://towardsdatascience.com/machine-learning-basics-part-1-a36d38c7916
- A quick recap of ML…

Supervised Learning

Supervised Learning

Learn a function that maps an input to an output based on examples of input-output pairs.



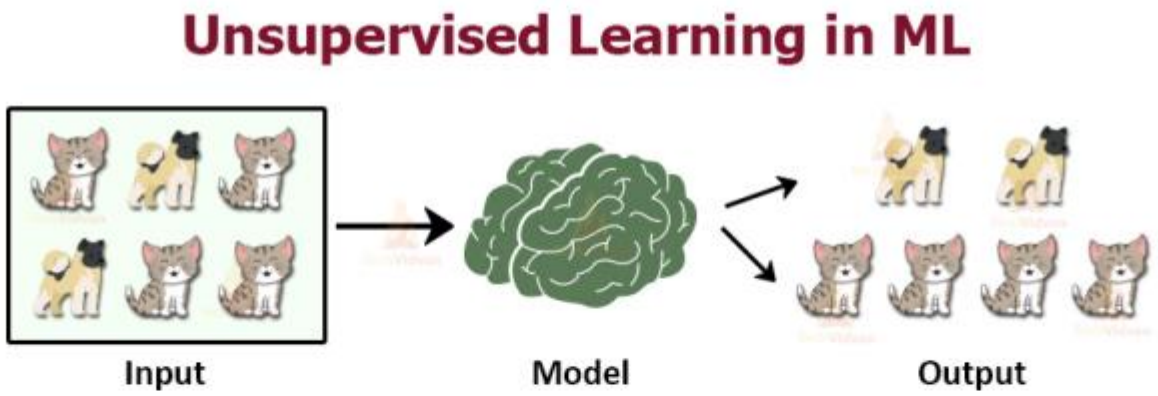
- Classification is used to predict discrete values (class labels).
- Regression is used to predict continuous values

Credits to Yunwen Lei

Unsupervised Learning

Unsupervised Learning

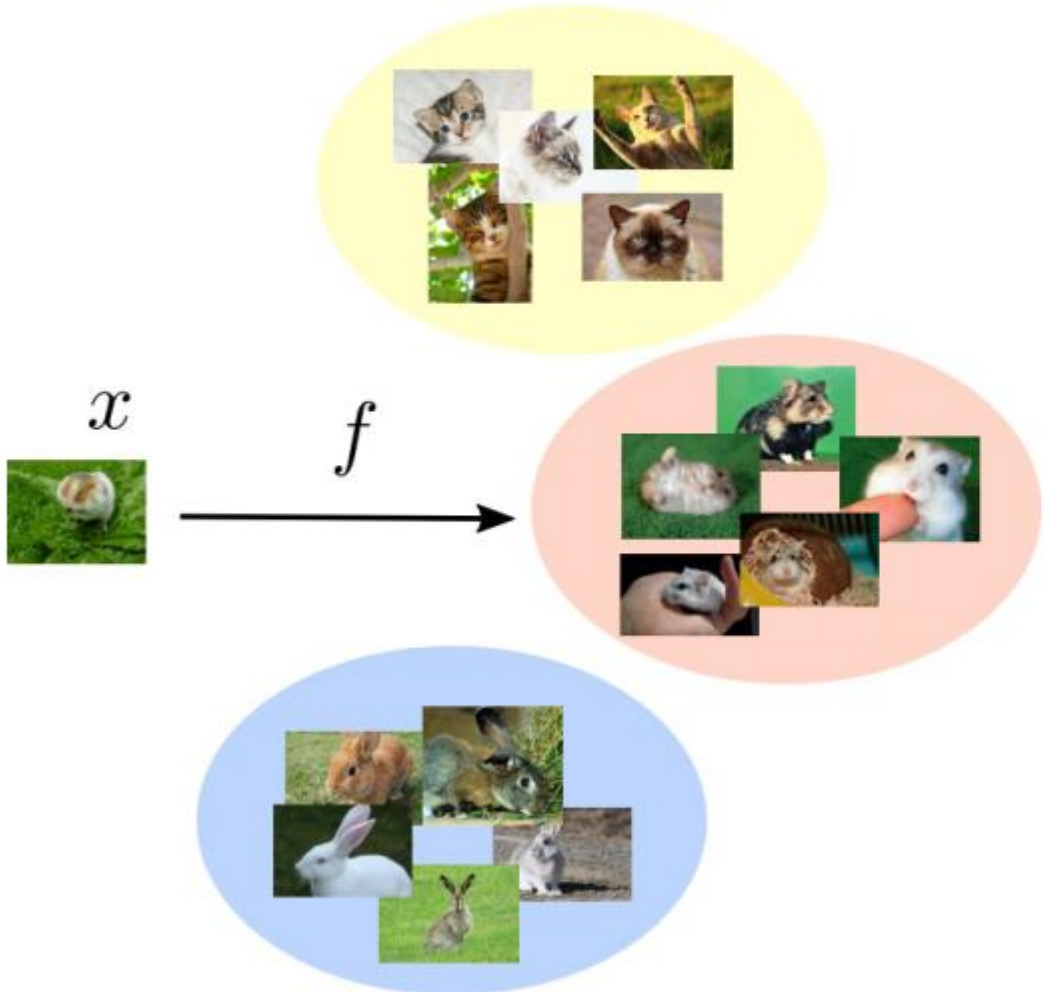
Learn interesting patterns from dataset with no labels: $\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(n)}$



- **Clustering** algorithm tries to detect similar groups.
- **Dimensionality reduction** tries to simplify the data without losing too much information.

Credits to Yunwen Lei

Classification



- Construct a function

$$f : \mathbb{R}^d \mapsto \{1, \ldots, k\}$$

such that if an object with features $\mathbf{x} \in \mathbb{R}^d$ belongs to class y then

$$f(\mathbf{x}) = y$$

- Alternatively, construct a function which given features returns the probability of each class

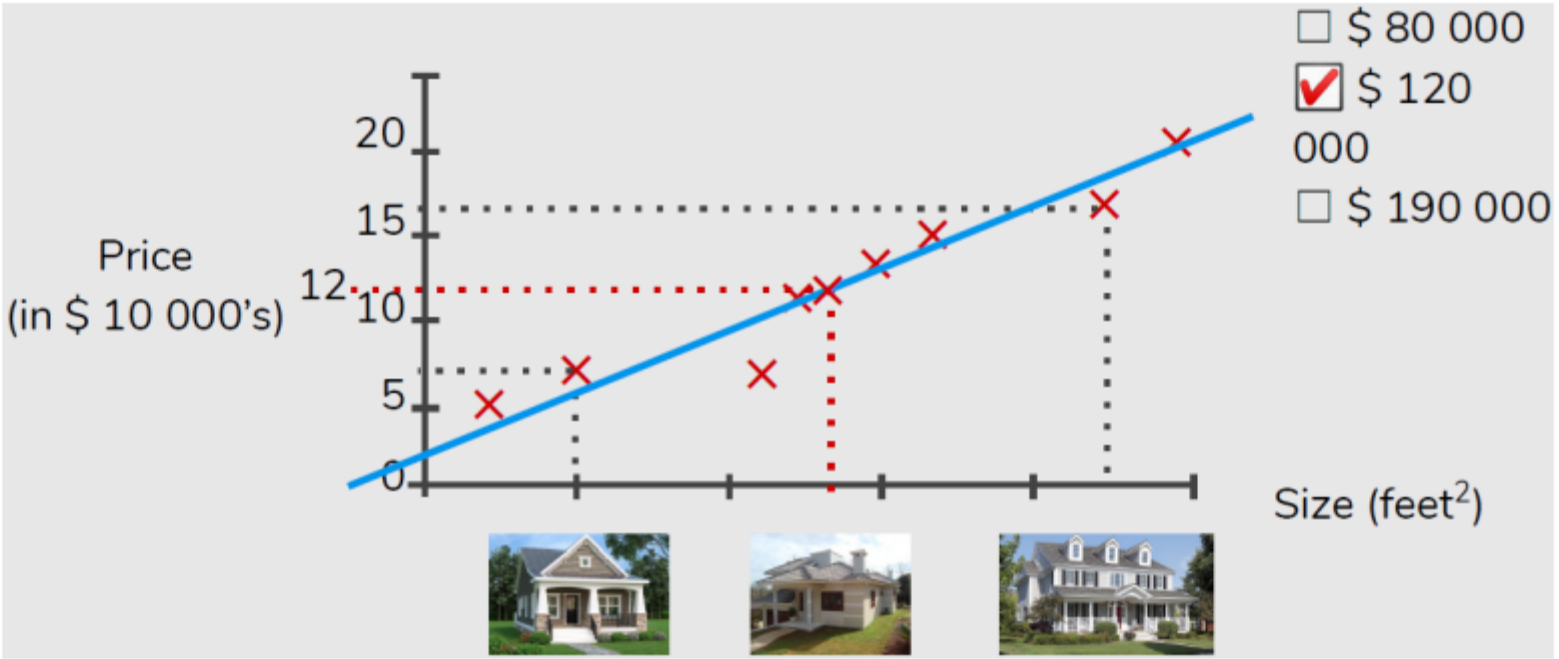
Credits to Yunwen Lei

Regression

- Predict a numerical output given some input, i.e., a function

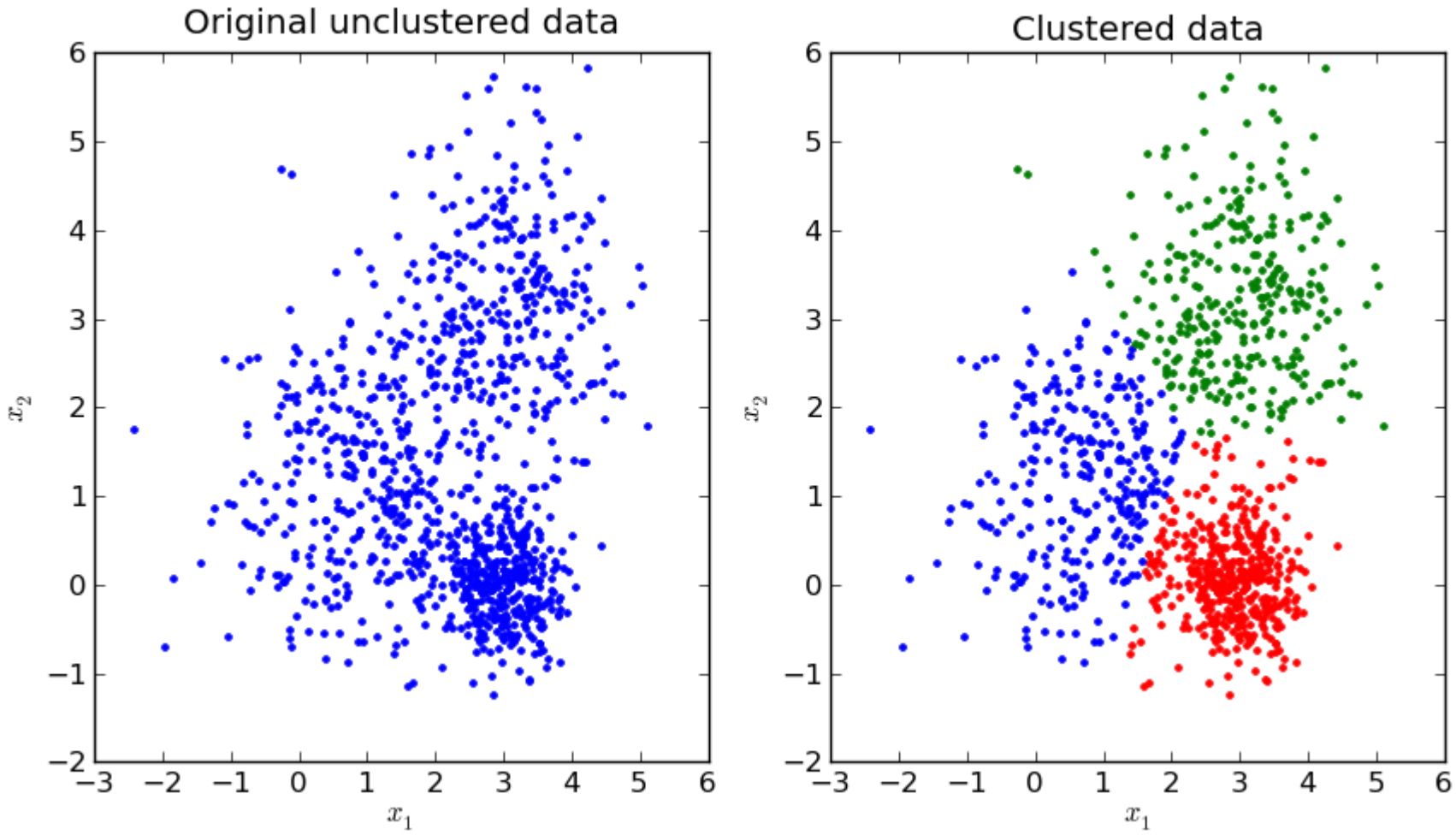
$$f : \mathbb{R}^d \mapsto \mathbb{R}.$$

- Example: house price prediction
  - **Input**: Information of House (living size, lot size, location, # floors)
  - **Output**: Price



Credits to Yunwen Lei

Clustering



Credits: https://towardsdatascience.com/k-means-data-clustering-bce3335d2203

ML: Training and Testing

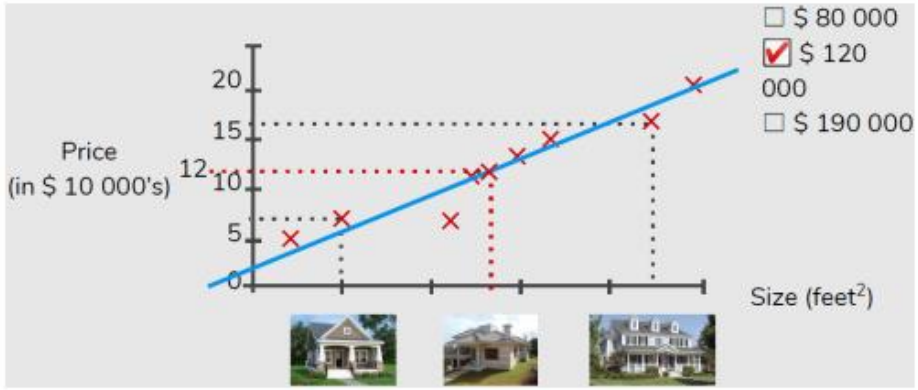
- **Input**: Information of House (living size, lot size, location, # floors)
- **Output**: Price
- **Aim**: find a model to predict price based on feature information of house
- **Training dataset**: a sequence of (features, price) pairs

$ 70,000

$ 160,000

- **Prediction/testing**: given information of new house, predict its price?

???



- **Performance**: difference between predicted price and true price

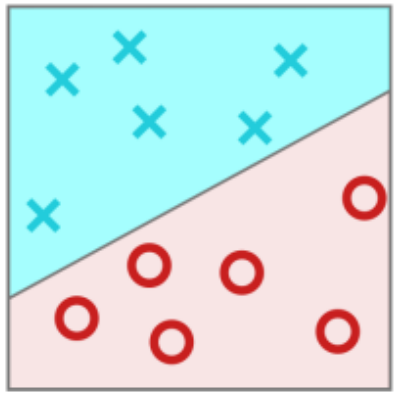
Credits to Yunwen Lei

Training and Testing

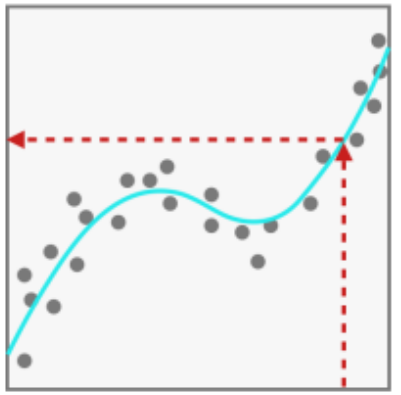
- **Training dataset**: a dataset that contains n samples

$$(\underbrace{\mathbf{x}^{(1)}}_{\text{input}}, \underbrace{y^{(1)}}_{\text{output}}), (\underbrace{\mathbf{x}^{(2)}}_{\text{input}}, \underbrace{y^{(2)}}_{\text{output}}), \ldots, (\underbrace{\mathbf{x}^{(n)}}_{\text{input}}, \underbrace{y^{(n)}}_{\text{output}})$$

- Classification: $y \in \{-1, +1\}$
  $+1$ means positive examples
  -1 means negative examples
- Regression: $y \in \mathbb{R}$



Classification



Regression

- Aim to find a function $f : \mathcal{X} \mapsto \mathcal{Y}$ such that

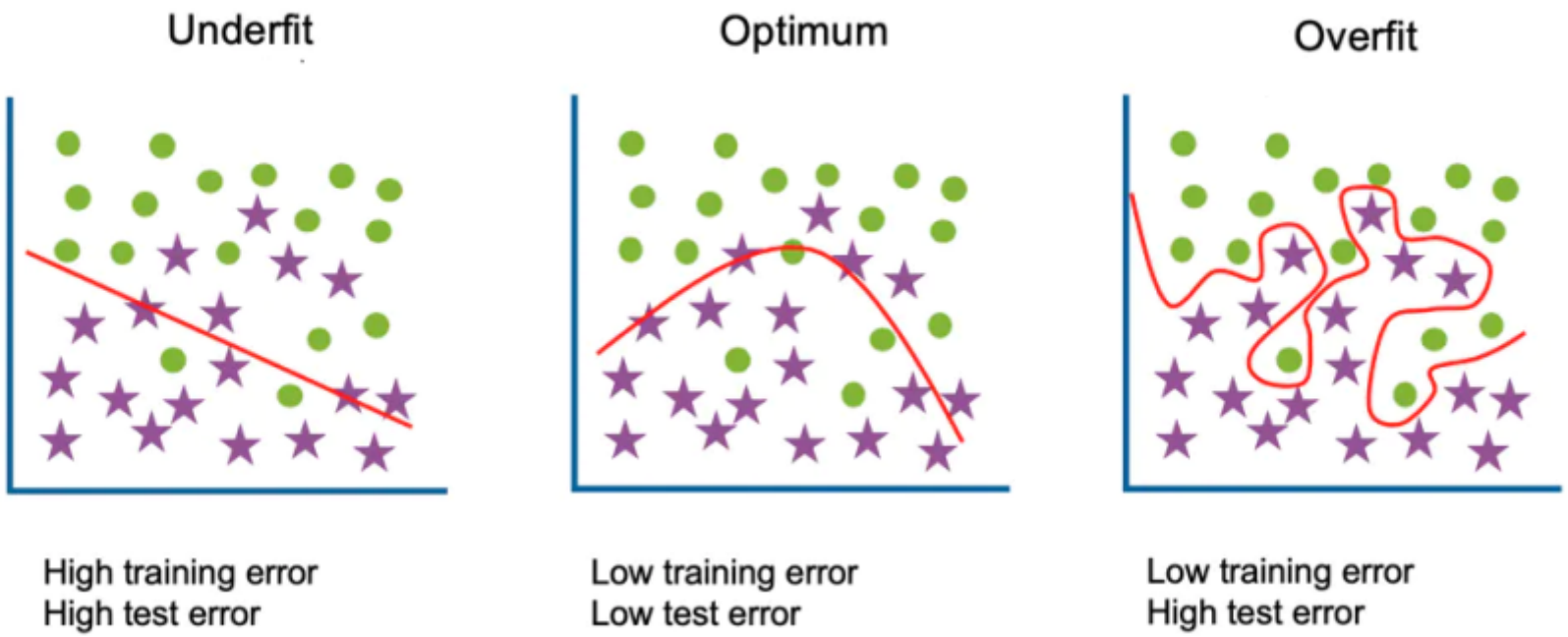
$$y \approx f(\mathbf{x})$$

- **Prediction**: given a new input $\mathbf{x}$, use f to do prediction
  - if a house has x square feet, predicting its price?

Credits to Yunwen Lei

Underfitting and Overfitting

- Loosely speaking, we say a model underfits when
  - training performance is poor
- We say a model overfits when
  - training performance is good but
  - test performance is poor



Credits to Yunwen Lei

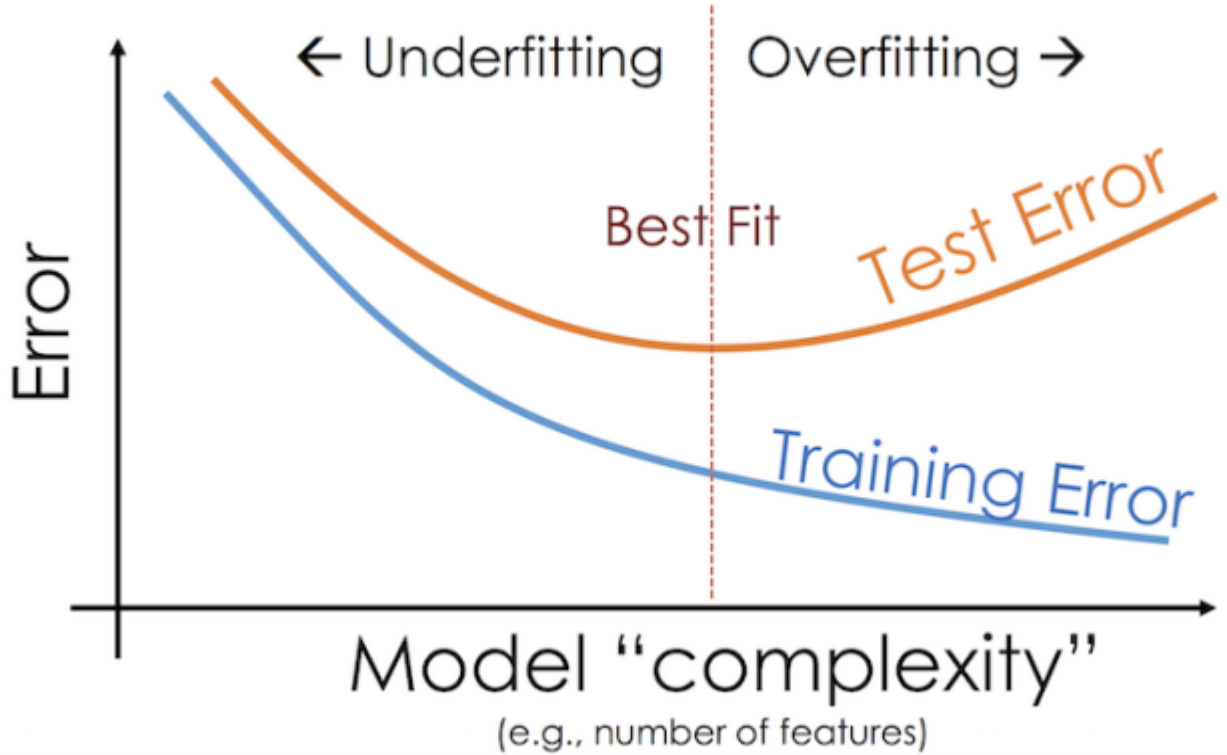
Overfitting the Training Data

- Overfitting means that the model performs well on the training data but it does not generalize to testing data
- It happens when the model is too complex relative to the amount and noisiness of the training data.

Credits to Yunwen Lei

Underfitting and Overfitting

- In general, the training error decreases as we add complexity to our model with additional features or more complex prediction mechanisms.
- The test error, on the other hand, decreases up to a certain amount of complexity then increases again as the model overfits the training set.



Credits to Yunwen Lei

Research introduction (1 of 3)

- What's research?
- What's a research paper?
- How's a paper published?
  - e.g., who decides that a paper is worth publishing?
  - Peer review process
- Why people write a research paper?
- Why study a research paper?

Research introduction (2 of 3)

- Where to find a research paper?
  - e.g., google scholar, arxiv, paperswithcode
- What's the difference between a conference and a journal paper?
- What's a good research paper?
  - e.g., where published? how many citations? which research group/university? what contributions made?

Research introduction (3 of 3)

- What's SOTA?
  - State-of-the-art or cutting edge
  - Forefront of developments

- How to understand SOTA?
  - Read & understand lots of papers
  - Look and read relevant good-quality review papers

How to read a research paper?

1. http://ccr.sigcomm.org/online/files/p83-keshavA.pdf
2. https://www.elsevier.com/connect/infographic-how-to-read-a-scientific-paper
3. https://www.huffpost.com/entry/how-to-read-and-understand-a-scientific-paper_b_5501628
4. https://blogs.lse.ac.uk/impactofsocialsciences/2016/05/09/how-to-read-and-understand-a-scientific-paper-a-guide-for-non-scientists/

Learning outcomes (repeated)

1. Demonstrate an understanding and appreciation of recent advances in AI/ML/DS

2. Make effective oral and written presentations

3. Engage effectively in discussions about recent research

- Let's discuss each one of them

MUST DO before week 2 and week 3 lectures

- Watch week 1 pre-lecture recommended videos
- New to AI/ML? Begin to read more about introductory ML
- Watch & read week 3's material on Canvas (to be released)
  - Week 2 material on "Vision-based Pose Estimation" is an optional watch (not formally applicable to Dubai)
- Study at least a couple of research paper
  - Practice/learn where to search
  - Practice/learn what to search
  - Practice/learn to select suitable/relevant search results
  - Read at least one research paper, following the first pass approach

Reading group

- Let’s read a paper… use first pass approach
  - Relatively easy difficulty papers
    - Computing Machines and Intelligence, Turing, 1950
    - The AI writing on the wall
  - Relatively medium difficulty papers
    - Weakly Supervised Captioning of Ultrasound Images
    - Simultaneous Semantic and Instance Segmentation for Colon Nuclei Identification and Counting
  - Relatively high difficulty paper
    - Development of metaverse for intelligent healthcare
- (In 20-30 minutes) Discuss the paper in your groups…