EECE 5644: Introduction to Machine Learning & Pattern Recognition Course Overview

Mark Zolotas

E-mail: m.zolotas@northeastern.edu
Webpage: https://coe.northeastern.edu/people/zolotas-mark/

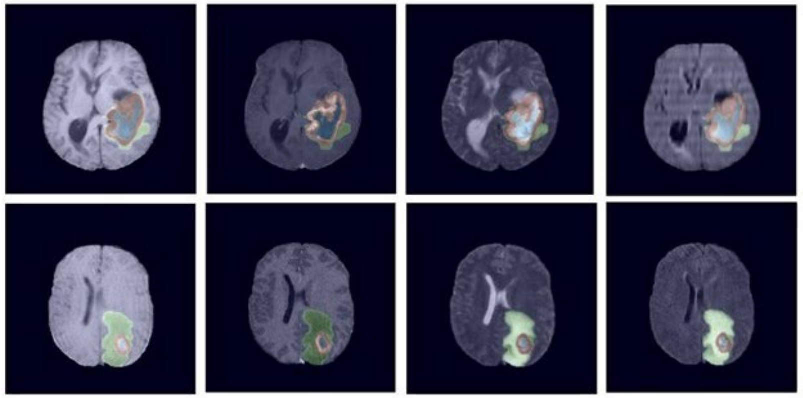
Personal Introduction

- **Imperial College London**, 2012-2020
  - Master's in Electronic & Information Engineering
  - Focus on robotics and machine learning
  - PhD from Personal Robotics Lab

- **NEU RIVeR Lab** (since Jan 2021)
  - Postdoctoral Researcher on teleoperation of robot arms
  - Personal approach involves a blend of robot control, machine learning and mixed reality

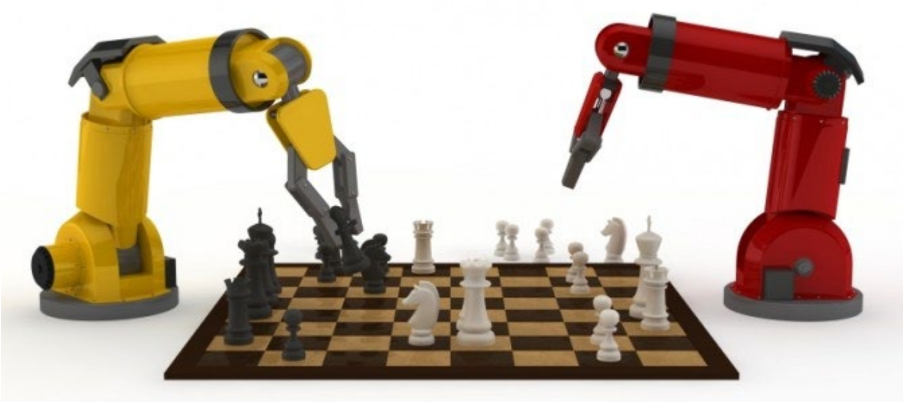
- No prof Zolotas, prof Mark, etc. please just call me Mark!

Why Machine Learning?



Brain Tumor Segmentation

Product Recommendation

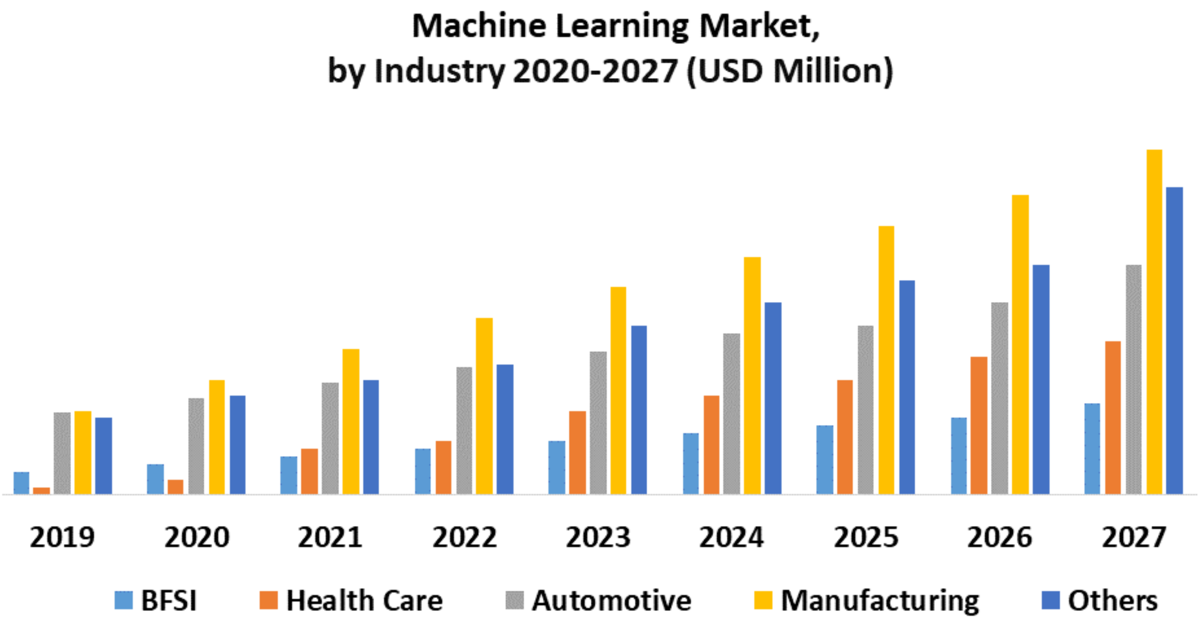


Robotics
observe→learn→act
repeat

Sources: InfoQ (Netflix), Nvidia (brain tumor segmentation), Dataaspirant (fraud detection), Robotnor (robotics)

Why… STILL Machine Learning?

- AI and data specialists **most sought for** roles in 2022 (Business Insider)
- In the US, machine learning roles continue to **grow** annually by **74%** (LinkedIn), or a total estimate of **44%** between **2017-2024** (Forbes)
- *Example:* **COVID-19 pandemic** to understand patterns in spread of virus or predict patient responses to different treatment plans (ScienceDirect)



Future of Many Sectors

Intelligent Transportation: route optimization, autonomous vehicles, improved safety & efficiency

Healthcare: personalized medicine, predictive diagnosis and treatment, "fully" robotic surgery

What Drives the Hype?

- **Ever-increasing data:** 2 to 97 zetabytes (2^{70} bytes!!!) from 2010 to 2022
- **Reduced storage costs**, can store terabytes cheaply
- **Improvements in compute** (CPUs, GPUs, …, TPUs)
- Better cooling solutions
- Other enhancements (research, education, ethics, etc.)
- BUT beware:
  - ❖ Learning algorithms are only as good as your **assumptions** and **data quality**
  - ❖ **NOT** every problem **needs** machine learning

What is Machine Learning?

Def. 1 (Arthur Samuel 1959):

"Field of study that gives computers the ability to learn without being explicitly programmed."

Def. 2 (Tom Mitchell 1997):

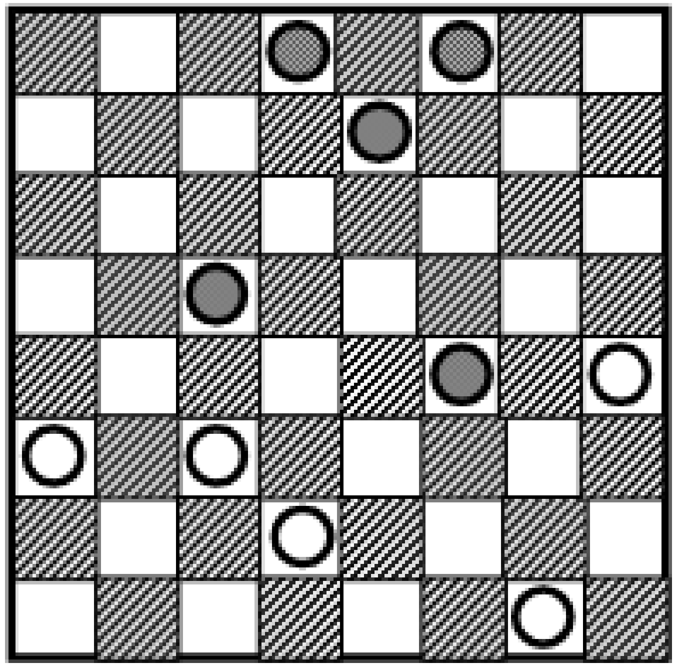
"A computer program is said to **learn** from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E."

Example: Learn to play checkers

T: *play checkers and win*

P: *% of games won*

E: *play against self and others*



Relationship to Other Fields

- Learning ↔ Intelligence
  - *Intelligence is the ability to learn/deal with new or trying situations*
- Machine Learning ↔ Artificial Intelligence (AI)
  - *AI is the science of making machines do things that would require intelligence if done by men* [Marvin Minsky 1986]
  - *Machine learning is an area of AI concerned with the development of techniques which allow machines to learn*
- Data Science ↔ Machine Learning
  - *Data science applies techniques from machine learning but focuses on data visualization and presentation (insights/analysis)*
- Pattern Recognition ↔ Machine Learning
  - *Machine learning provides a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty (making an 'action')* [Kevin Murphy 2012]

Course Objectives

- Introductory course on machine learning, so we will cover a **range of algorithms** for a broad understanding of the topic
- Enable students to learn **where and how to apply machine learning** algorithms, as well as why they work in some settings but not others
- Equip students with the foundations for a future in machine learning through exposure to **theory** and **hands-on experience**
- Present students with an opportunity to **dive deeper** into a subtopic
  - Write a technical paper in the chosen area
  - Formulate the problem and implement it programmatically
  - Present the results

Understanding

Use-cases

Theory/Practice

Depth

Prerequisites

Probability EECE 3468/MATH3081 or equivalent for undergraduates, EECE 7204/DS5020 or equivalent for graduate students

Linear Algebra

https://cs229.stanford.edu/section/cs229-linalg.pdf

https://www.math.uwaterloo.ca/~hwolkowi/matrixcookbook.pdf

Programming: Must be a self-sufficient programmer

(**Python**, MATLAB, C/C++, R, etc.)

~60% of machine learning developers use Python…

Course Evaluation

- **Take-home Assignments (3*25%):** Four take-home assignments will account for 75% of your grade, with the three best scores taken
- Assignments will be handed out **weekly** on Thursdays/Fridays, and then due the following Monday (9-10 days later)
- **Project (remaining 25%):** Teams of 2-3 people (strictly) will work on a machine learning problem of their choosing; writing code, producing, reporting and presenting the results
- Project will occupy **approx. last 2 weeks** of the course – more info later

Course Resources

- **Lectures:** Distributed on a week-by-week basis
- **Code:** https://github.com/mazrk7/EECE5644_IntroMLPR_LectureCode
- **Canvas:** Will contain lecture slides, an updated timetable for the course and all handouts, e.g., for homework
- **Homework & Project Report** submitted on Canvas with all deadlines of 11:59PM EDT on the due date
- **Microsoft Teams Channel:** General class updates, discussions & questions about homework, course material, code, etc… **Do NOT directly message me on Teams**
- **Teaching Assistant (TA):** Paul Ghanem ghanem.p@northeastern.edu
- **Office Hours:** Mon, Wed 2-3:30pm with Paul and Fri 2-4:00pm with me

Supplementary Material

Textbooks NOT required but recommended as a formal reference:

- Kevin P. Murphy, *Machine Learning: A Probabilistic Perspective*, MIT Press 2012
- Kevin P. Murphy updated book in draft state: *Probabilistic Machine Learning: An Introduction*, 2022 ← Most recommended
- Christopher M. Bishop, *Pattern Recognition and Machine Learning*, Springer 2006
- R. O. Duda, P. E. Hart, D. Stork, *Pattern Classification*, *2nd Ed*, Wiley and Sons, 2001
- I. Goodfellow, Y. Bengio, A. Courville, *Deep Learning*, MIT Press, 2016

Tentative Course Outline (Wks. 1-2)

Topics	Dates	Assignments	Additional Reading
Course Overview Machine Learning Basics	07/05	**Optional Homework 0** released on Canvas on 07/08 but please do NOT submit on Canvas	Chpt. 1 Murphy 2012
Foundations: Linear Algebra, Probability, Numerical Optimization (Gradient Descent), Regression	07/06-11		Stanford LA Review Stanford Prob. Review Chpt. 8 Murphy 2022
Quick Python Tutorial	07/12	**Homework 1** released on Canvas on 07/15 **Due 07/25**	N/A
Linear Classifier Design, Linear Discriminant Analysis and Principal Component Analysis (PCA)	07/13-14		Chpts. 9.2 & 20.1 Murphy 2022
Bayesian Decision Theory: Empirical Risk Min, Max Likelihood (ML), Max a Posteriori	07/14-15		Chpt. 2 Duda & Hart 2001 Deniz Erdogmus Notes

Tentative Course Outline (Wks. 3-4)

Topics	Dates	Assignments	Additional Reading
Naïve Bayes Classifier & *Homework 0 Practice Lab*	07/18	**Homework 2** released on Canvas on 07/22 **Due 08/01**	N/A
Model Fitting/Training: Bayesian Parameter Estimation	07/19-20		Chpts. 4.1-4.3, 8.7.2-3 Murphy 2022
Logistic Regression	07/21		Chpt. 10 Murphy 2022
Model Selection: Hyperparameter Tuning, k-fold Cross-Validation	07/25	**Homework 3** released on Canvas on 07/29 **Due 08/08**	Chpts. 4.5, 5.2, 5.4.3 Murphy 2022
Regularization, Ridge and Lasso Regression	07/26		Chpts. 4.5, 11.1-11.4 Murphy 2022
Neural Networks: Multilayer Perceptrons & Backpropagation	07/27-28		Chpts. 13.1-13.5 Murphy 2022

Tentative Course Outline (Wks. 5-6*)

<table>
<tr><th>Topics</th><th>Dates</th><th>Assignments</th><th>Additional Reading</th></tr>
<tr><td>Support Vector Machines (SVMs)</td><td>08/01-02</td><td rowspan="3">Homework 4 released on Canvas on 08/05
Due 08/15</td><td>Burges Tutorial</td></tr>
<tr><td>Clustering: K-means, Gaussian Mixture Models (GMMs)</td><td>08/03</td><td>Chpt. 21 Murphy 2022</td></tr>
<tr><td>Labs/Recap/Interactive/Catch-up</td><td>08/04</td><td>N/A</td></tr>
<tr><td>More on Deep Learning (CNNs & RNNs)</td><td>08/08-09</td><td rowspan="2">Make sure project teams (2-3 ppl. strict) are fully formed by 08/12</td><td>Deep Learning Goodfellow et al. 2016
http://d2l.ai/</td></tr>
<tr><td>Ensemble Methods: Decision Trees, Boosting & Bagging</td><td>08/09-10</td><td>Chpt. 18 Murphy 2022</td></tr>
</table>

Tentative Course Outline (Wks. 6*-8)

<table>
<tr><th>Topics</th><th>Dates</th><th>Assignments</th><th>Additional Reading</th></tr>
<tr><td>Project + Practical Tips</td><td>08/11</td><td rowspan="5">Final Project Reports & Code
Due 08/22

Presentations on 08/22-23 in normal lecture hours and office hours depending on no. of groups</td><td>N/A</td></tr>
<tr><td>Dimensionality Reduction & Representation Learning (Autoencoders)</td><td>08/15</td><td>Chpt. 20
Murphy 2022</td></tr>
<tr><td>Model Predictive Control (MPC)</td><td>08/16-17</td><td>TBD</td></tr>
<tr><td>Gaussian Processes</td><td>08/18</td><td>TBD</td></tr>
<tr><td>Project Presentations</td><td>08/22-23</td><td>N/A</td></tr>
</table>

Questions?

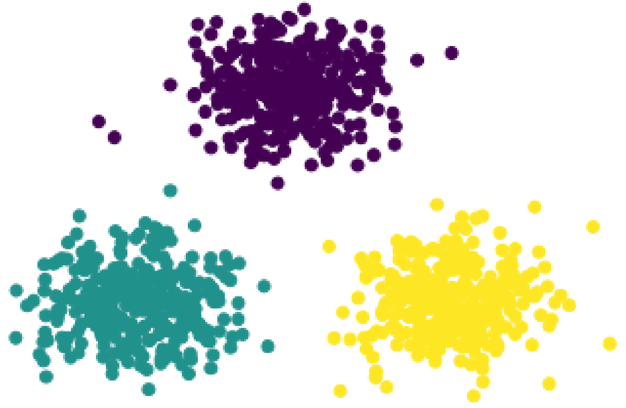
Course Layout

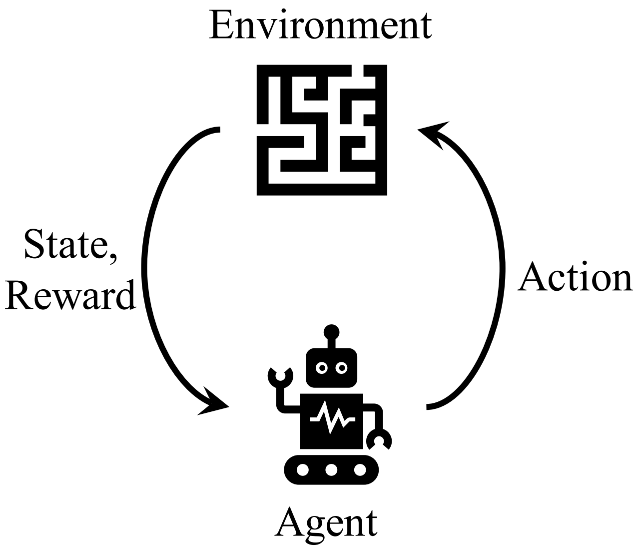
Examination

Resources

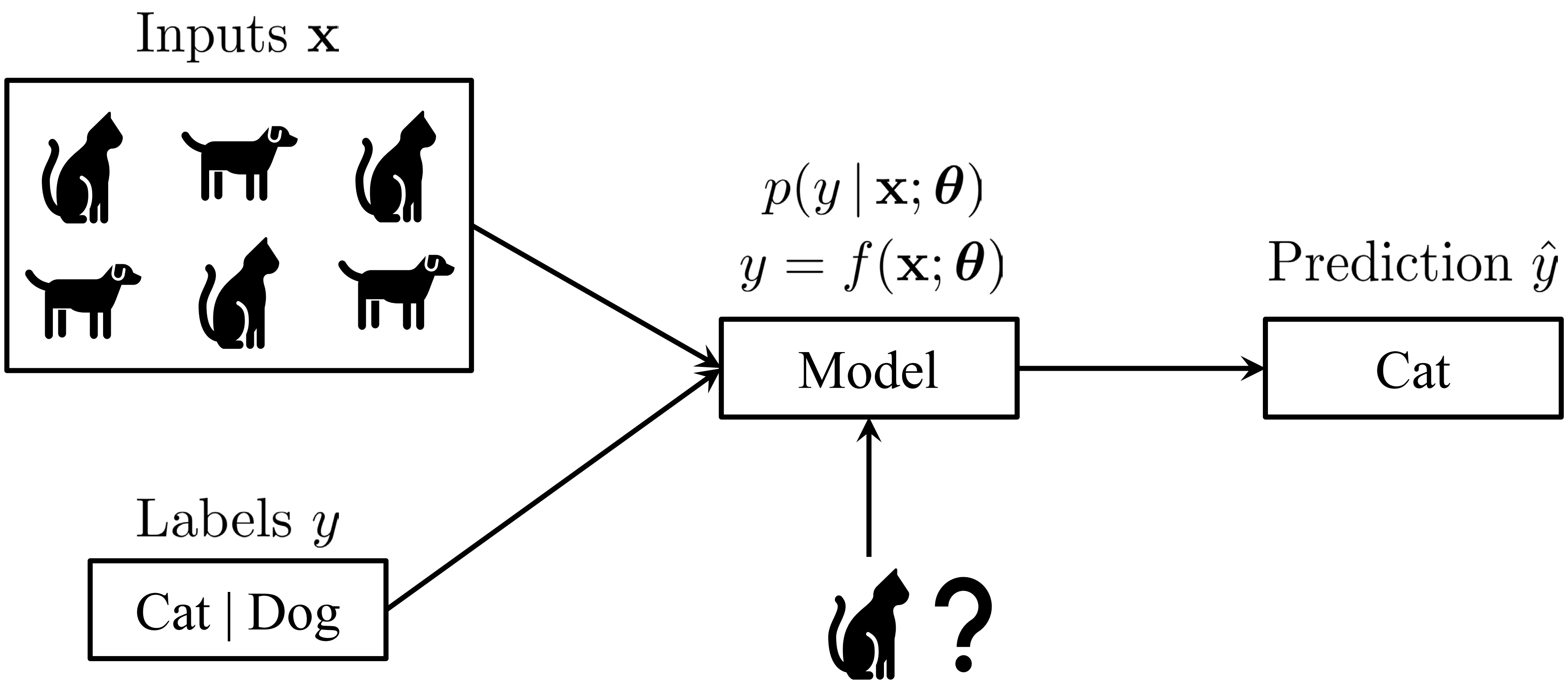
Types of Machine Learning

- **Supervised Learning:** Train or "teach" an algorithm using input-output pairs (labelled/categorized data)
  - Classification
  - Regression
- **Unsupervised Learning:** No feedback, "make sense" of structure in the data (*knowledge discovery*)
  - Clustering
  - Dimensionality Reduction (e.g., PCA)
  - Feature Learning (e.g., Autoencoders)
- **Reinforcement Learning:** Equip intelligent agents with reward-maximizing decision-making (action-taking)



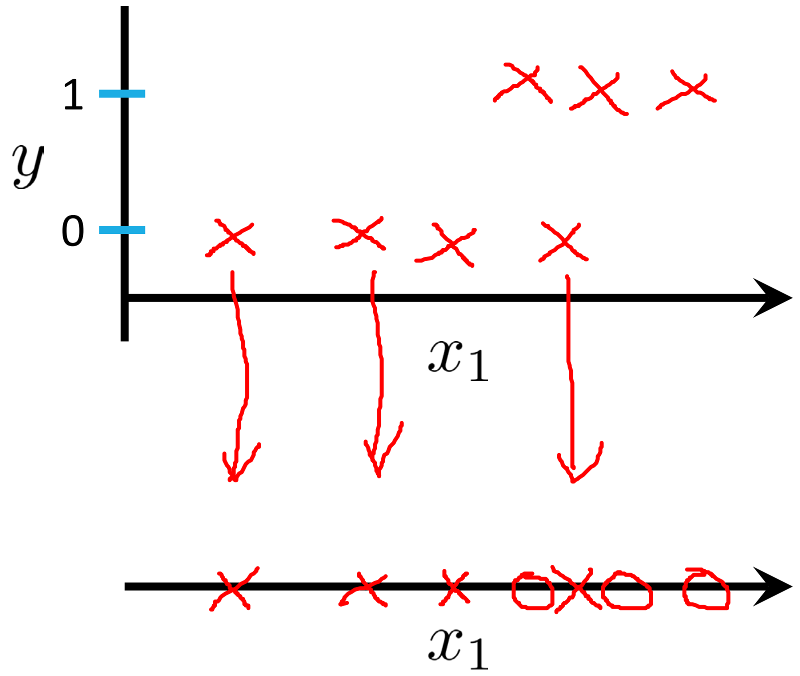


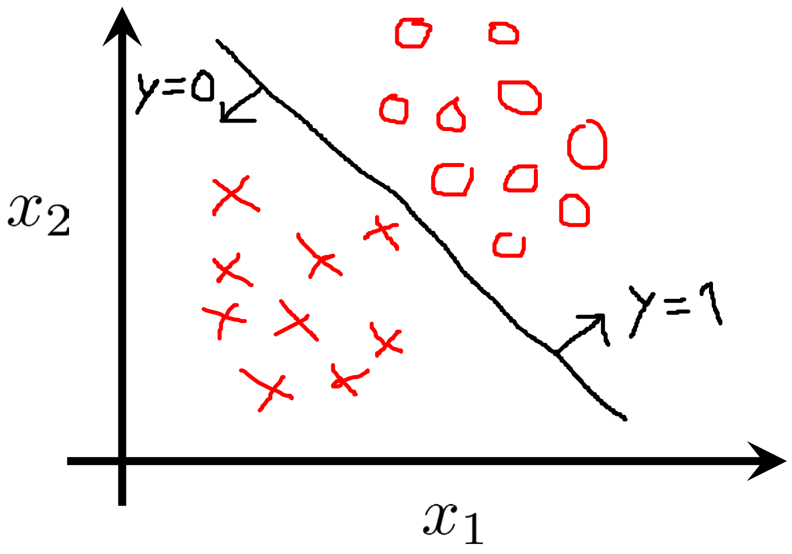
Supervised Learning



Classification

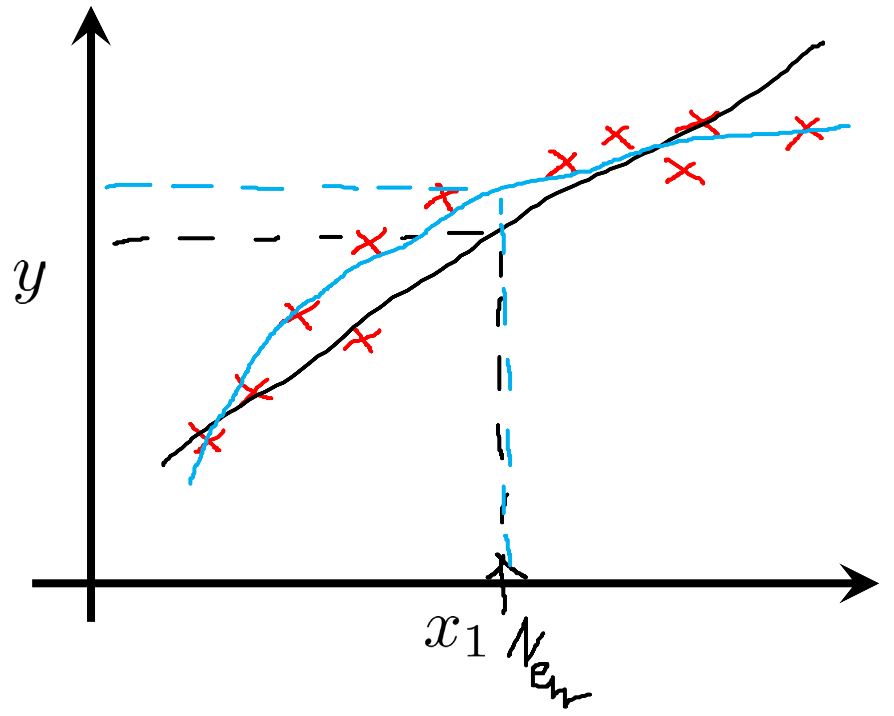
Let $\mathcal{D} = \{(\mathbf{x}^{(i)}, y^{(i)})\}_{i=1}^{N}$, N training samples
Inputs or **features** $\mathbf{x} \in \mathcal{X} = \mathbb{R}^n$
Classification: discrete valued outputs or **labels** $y \in \{1, \ldots, C\}$
Binary if $C = 2$, i.e. $y \in \{0, 1\}$, **multi-label** if $C > 2$





Regression

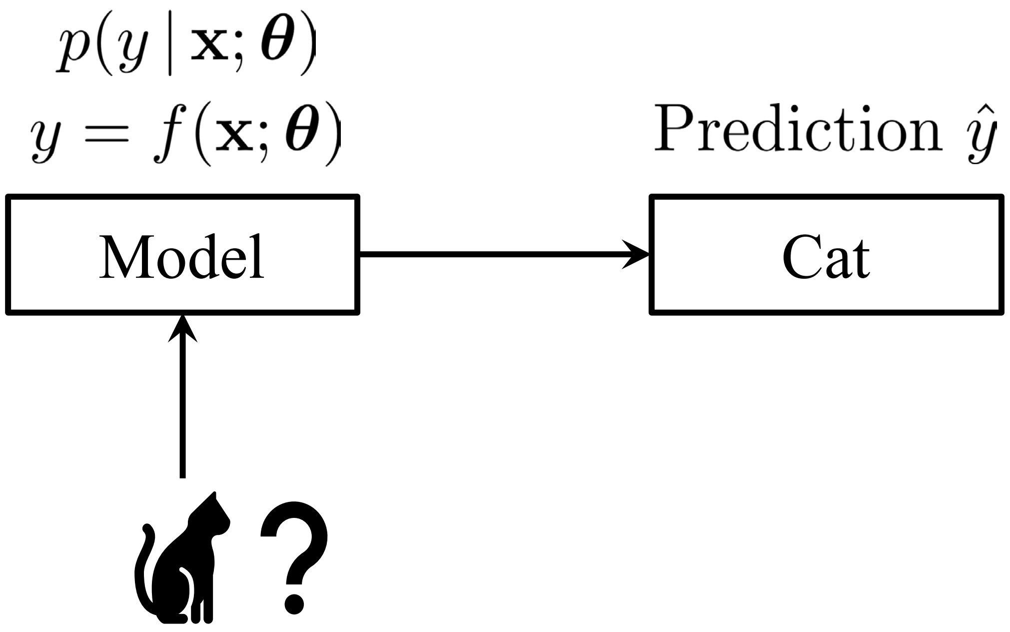
Regression: same as before, except **continuous** valued outputs
For i^{th} feature vector $\mathbf{x}^{(i)} \in \mathbb{R}^n$, there is a real-valued response $y^{(i)} \in \mathbb{R}$



Generalization

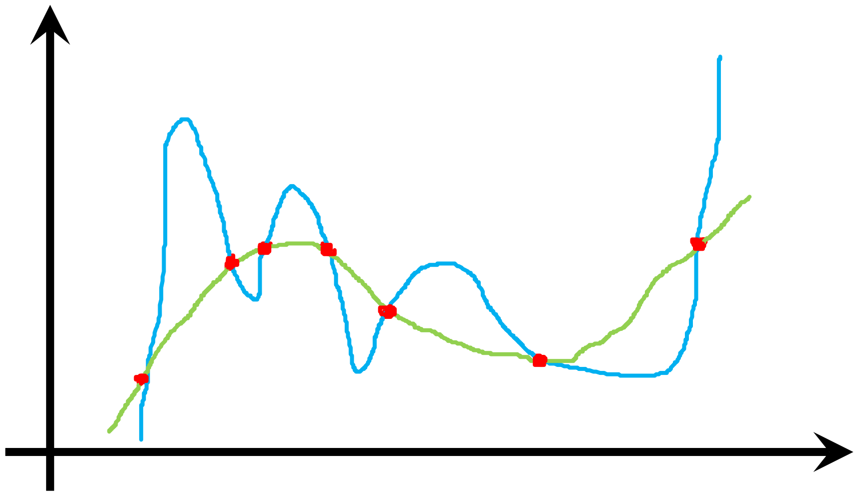
In both classification and regression; wish to predict on *future/unseen* data

Aim to **generalize** beyond our "training set"



Overfitting/Underfitting (1)

Def. Given a hypothesis space $\mathcal{H}$, a hypothesis $h \in \mathcal{H}$ is said to **overfit** the training data if there exists some alternative hypothesis $h' \in \mathcal{H}$, such that h has smaller error than h' over the training examples, but h' has a smaller error than h over the entire distribution of instances (Mitchell, 1997).



Red: Training set
Green: True target function
Blue: What we have learned (overfit)

Overfitting: Small error on training set, large on **test set**
Underfitting: Large **error** on both training and test sets

Overfitting/Underfitting (2)

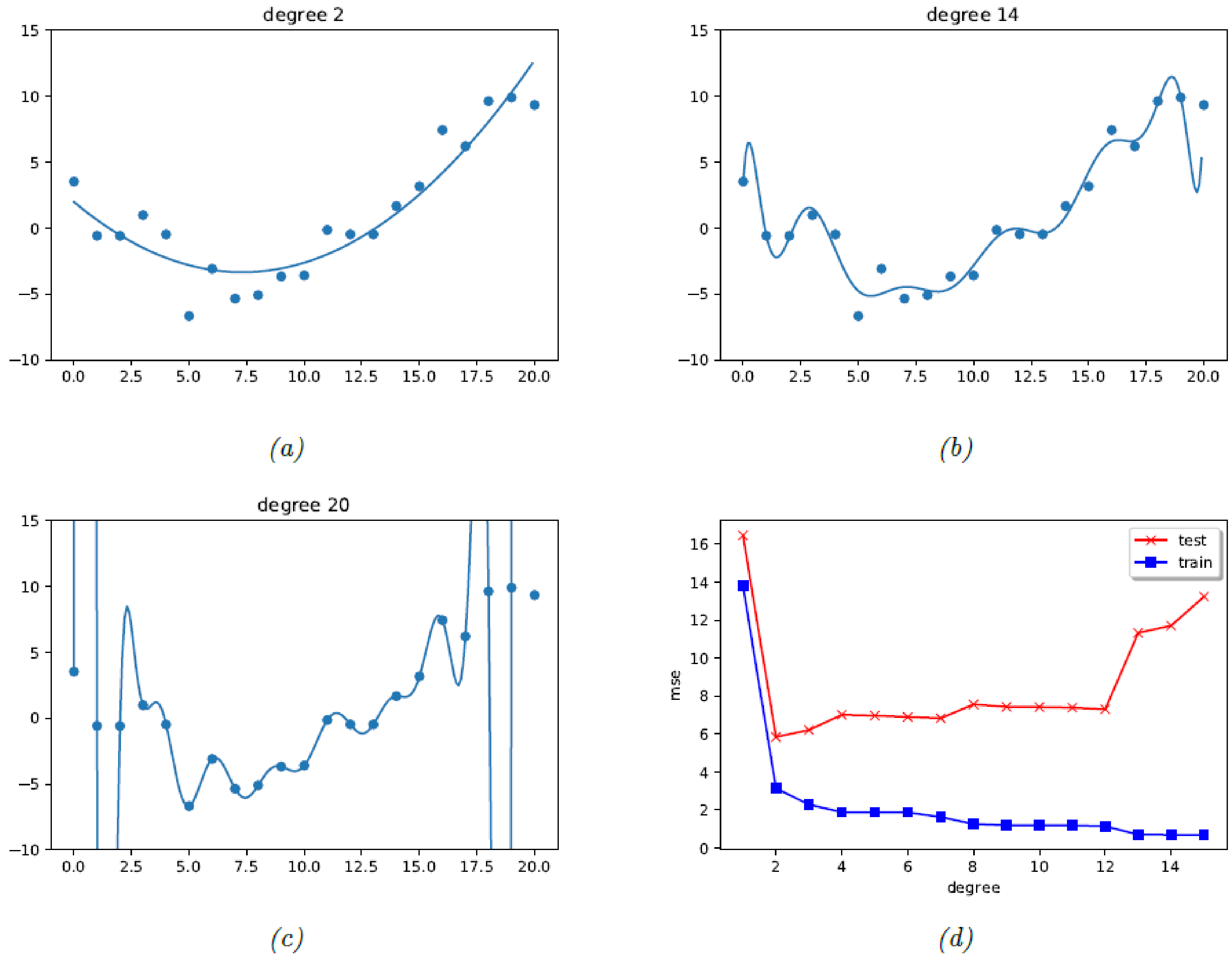
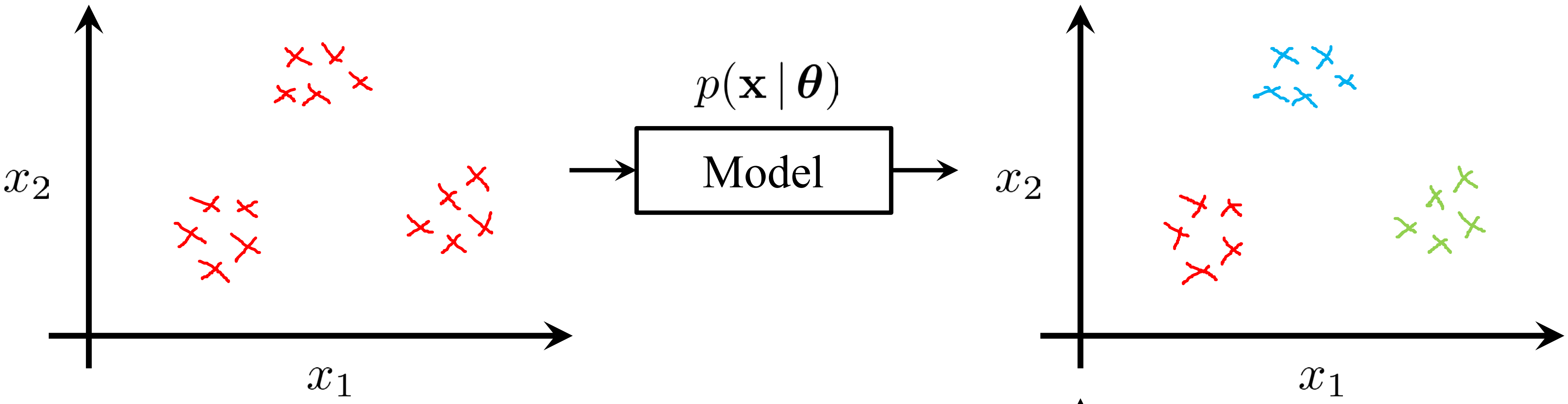


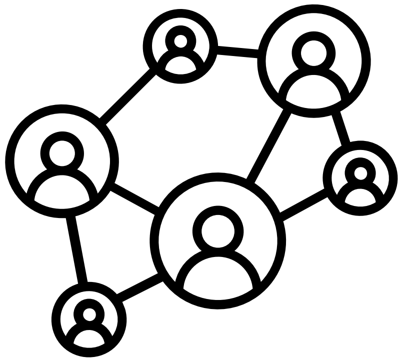
Figure 1.7: (a-c) Polynomials of degrees 2, 14 and 20 fit to 21 datapoints. (d) MSE vs degree.

Murphy, *"Probabilistic Machine Learning: An Introduction"*, 2022

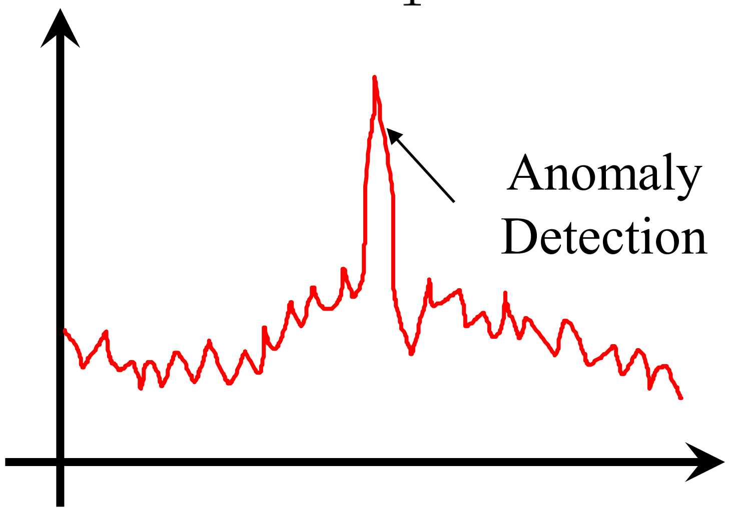
Unsupervised Learning



Social Network Analysis



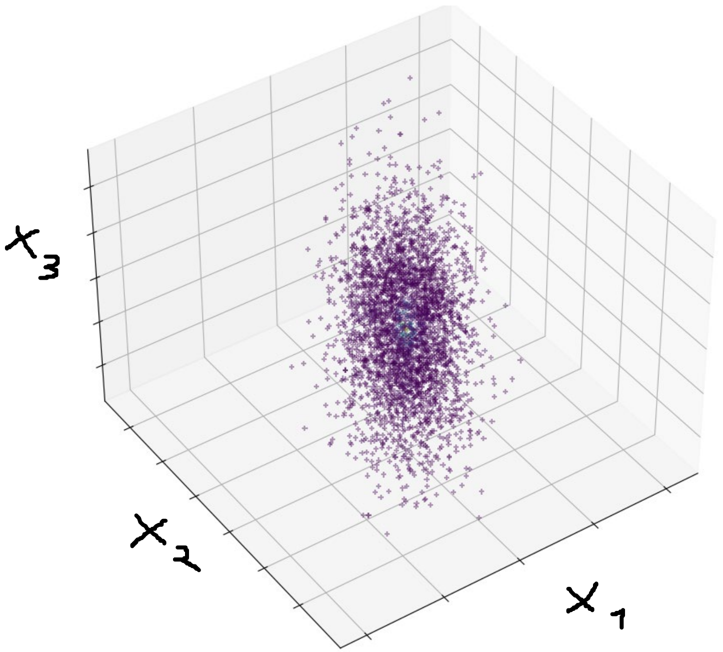
Customer Segmentation

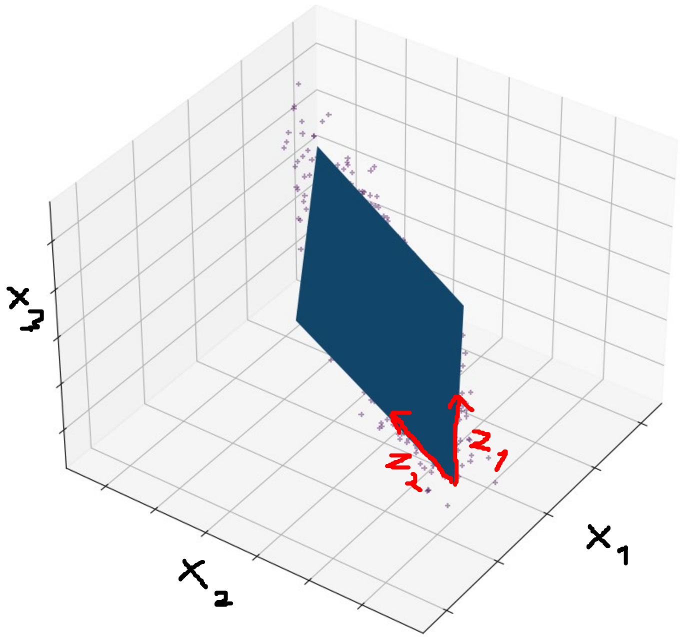


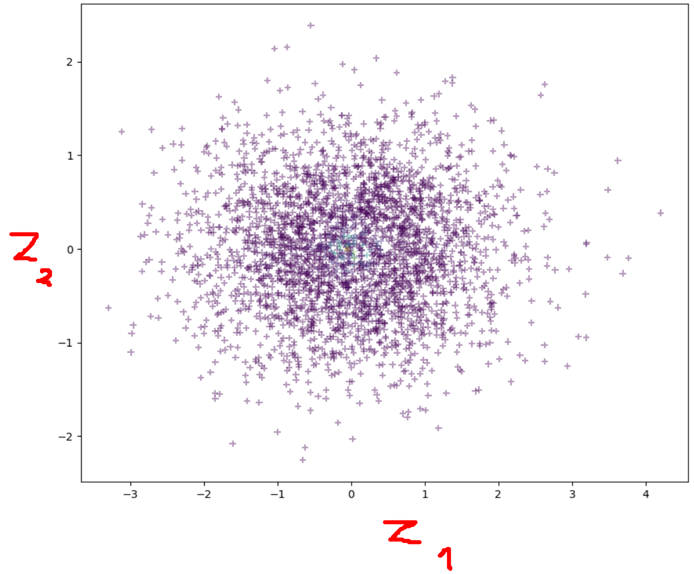
Dimensionality Reduction

$\mathbf{x} \in \mathbb{R}^d \rightarrow \mathbf{z} \in \mathbb{R}^m$ where $m < d$

$x \in \mathbb{R}^3$ $z \in \mathbb{R}^2$







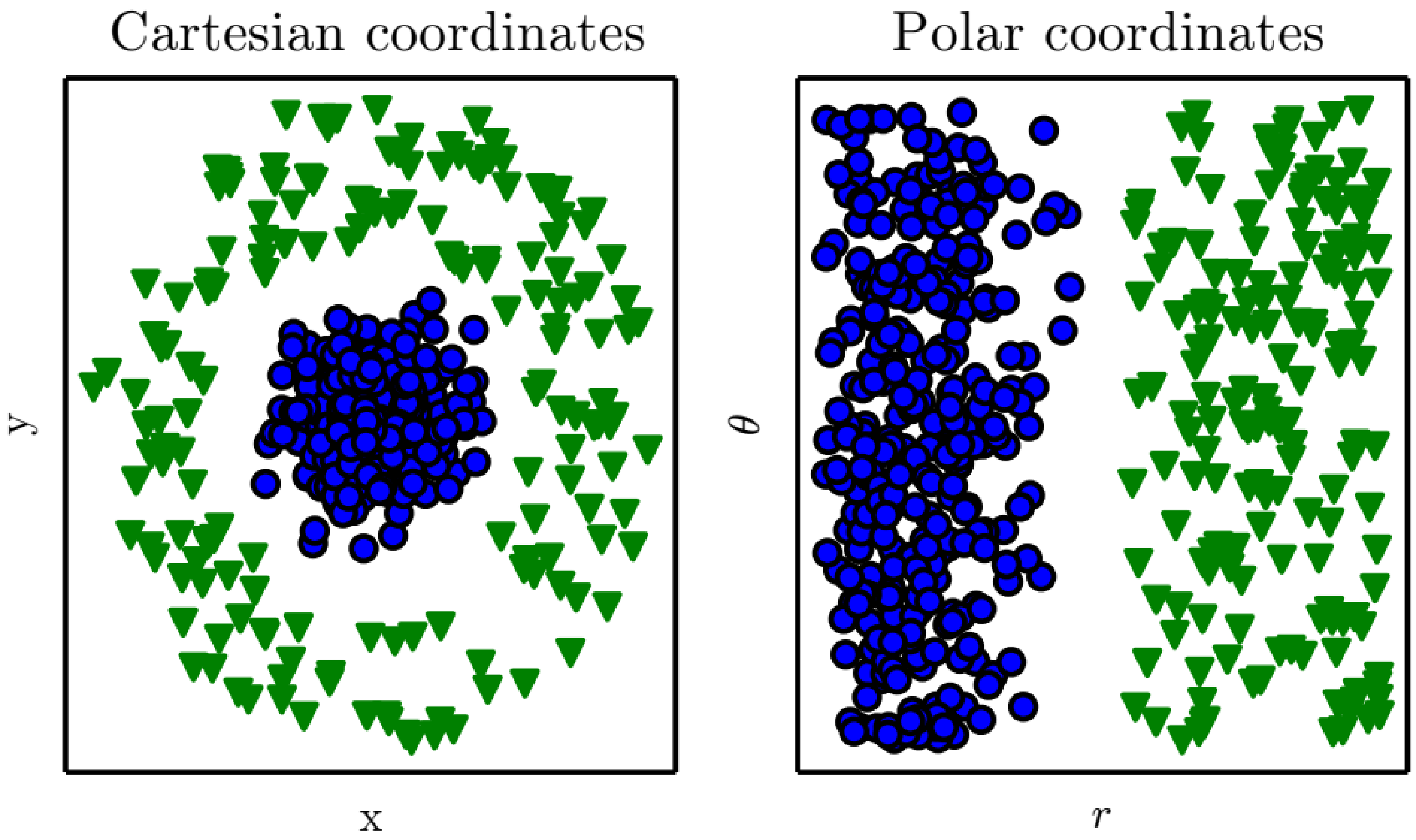
Useful for:

- Compression (remove redundancies)
- Visualization

Beware:

- Lossy transformation
- Prevent overfitting?

Representation Matters



Goodfellow et al., *"Deep Learning"*, 2016

Deep Learning

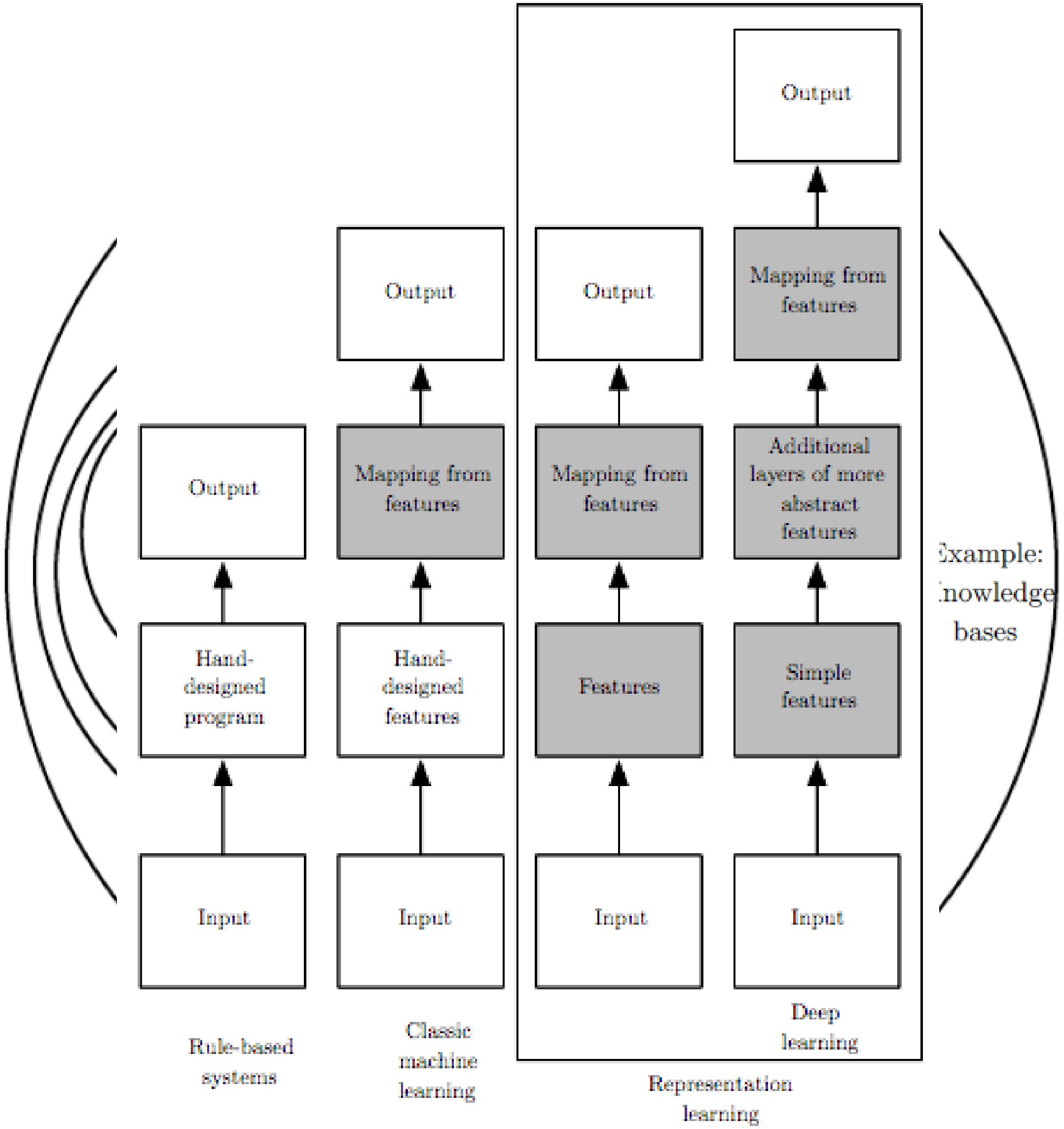


Figure 1.4: A Venn diagram showing how deep learning is a kind of representation learning, which is in turn a kind of machine learning, which is used for many but not all approaches to AI. Each section of the Venn diagram includes an example of an AI technology.

Goodfellow et al., *"Deep Learning"*, 2016

Figure 1.5: Flowcharts showing how the different parts of an AI system relate to each other within different AI disciplines. Shaded boxes indicate components that are able to learn from data.

Goodfellow et al., *"Deep Learning"*, 2016

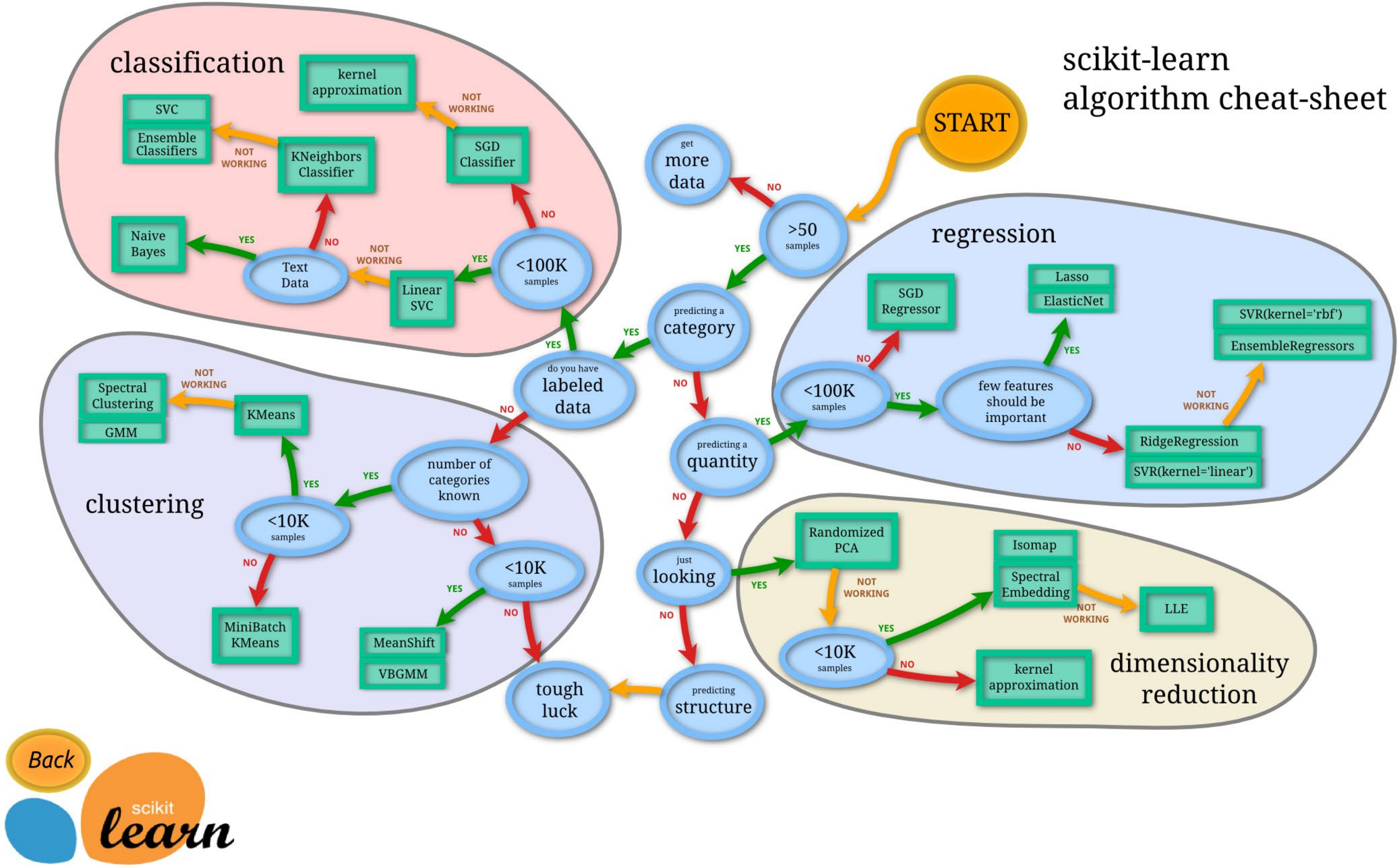
Quick Quiz

Are the following **supervised/unsupervised** applications of machine learning?

1) Given incoming labeled spam emails, learn an algorithm to identify spam: __S__
2) Provided with a labeled image dataset of social media photos, detect pixel regions where faces are present: __S__
3) Discover association rules between different Amazon products and customers: __U__
4) Labeling and ranking web pages on Google: __S+U__

Software Tools

https://towardsdatascience.com/best-python-libraries-for-machine-learning-and-deep-learning-b0bd40c7e8c


scikit-learn
algorithm cheat-sheet
START
>50 samples
get more data
predicting a category
do you have labeled data
<100K samples
Linear SVC
Text Data
Naive Bayes
KNeighbors Classifier
SVC
Ensemble Classifiers
SGD Classifier
kernel approximation
classification
clustering
number of categories known
<10K samples
KMeans
Spectral Clustering
GMM
MiniBatch KMeans
MeanShift
VBGMM
tough luck
predicting structure
predicting a quantity
just looking
regression
SGD Regressor
few features should be important
Lasso
ElasticNet
RidgeRegression
SVR(kernel='linear')
SVR(kernel='rbf')
EnsembleRegressors
dimensionality reduction
Randomized PCA
Isomap
Spectral Embedding
LLE
kernel approximation
YES
NO
NOT WORKING
Back
scikit
learn