Chapter 1: Machine Learning Paradigms

Advanced Topics in Statistical Machine Learning

Tom Rainforth

Hilary 2024

rainforth@stats.ox.ac.uk

Course Overview

Course Essentials

Who is the course for?

- MMath Part C, OMMS, MSc in Statistical Science, CDT and DPhil students, anyone else who wants to follow along

Prerequisites

- Only true prerequisite is basic statistics and maths; course should be self–contained
- SB2.2/SM4 Statistical Machine Learning will be very helpful: there is noticeable overlap but because it runs at the same time (for MSc) we will go over everything you need again

Course resources

- Everything should be available through Canvas `https://canvas.ox.ac.uk/courses/224564`

Course Structure

- Detailed slides provided for each lecture—these should be your primary point of reference
- Also separate set of course notes (~ 100 pages)—designed to supplement lectures, sometimes provides extra details or alternative viewpoints
- 4 problems sheets
- Recorded lectures from last year all already available on Canvas site, syllabus identical to last year

Assessment

- Written examination (undergrads and masters)
- D.Phil and CDT students wanting to do the course for credit instead will instead write a 12–page literature review on a topic from the course and need to make arrangements for their own marker (see course website for details)

Topics

1. Review of fundamentals (1 lecture)
2. Constrained optimization and SVMs (2 lectures)
3. Kernel methods (4 lectures)
4. Bayesian machine learning (1.5 lectures)
5. Gaussian processes (2 lectures)
6. Deep learning (3.5 lectures)
7. Latent variable models and variational approaches (2 lectures)

What is Machine Learning?

What is Machine Learning?

Arthur Samuel, 1959

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

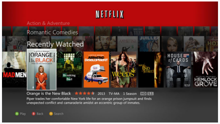
Tom Mitchell, 1997

Any computer program that improves its performance at some task through experience.

Kevin Murphy, 2012

To develop methods that can automatically detect patterns in data, and then to use the uncovered patterns to predict future data or other outcomes of interest.

What is Machine Learning?



recommender systems

machine translation

self-driving cars

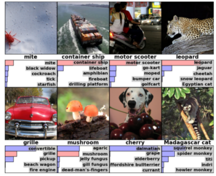
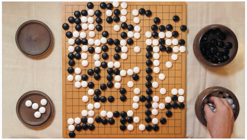


image recognition

DQN Atari games



AlphaGo

Learning from Data

- Machine learning is all about learning from data
- There is generally a focus on making predictions at unseen datapoints
- Starting point is typically a dataset—we can delineate approaches depending on type of dataset

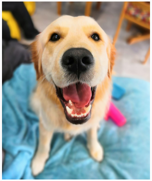
Supervised vs Unsupervised Machine Learning

Supervised Learning

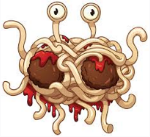
- We have access to a **labeled dataset** of input–output pairs: $\mathcal{D} = \{x_i, y_i\}_{i=1}^n$.
- Aim is typically to learn a **predictive model** f that takes an input $x \in \mathcal{X}$ and aims to predict its corresponding output $y \in \mathcal{Y}$.
- The hope is that these example pairs can be used to "teach" f how to accurately make predictions.
- Most common examples are classification and regression

Supervised Learning—Classification

Cat



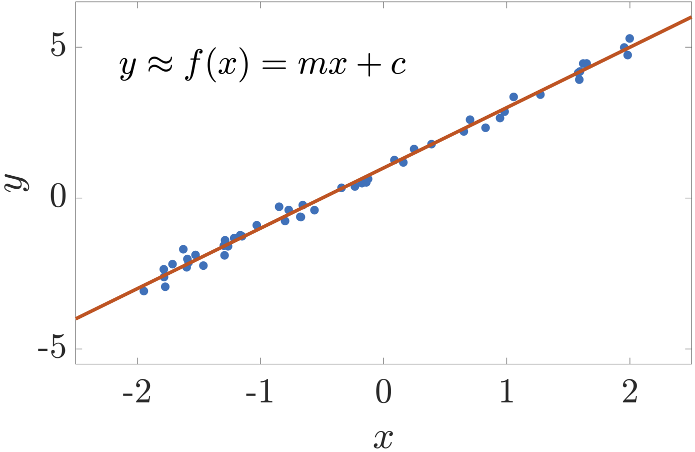
Dog



Flying
Spaghetti
Monster

Input x | Predictor $f(x)$ | Class label y

Supervised Learning—Regression



Supervised Learning

Training Data

Input Features

Outputs

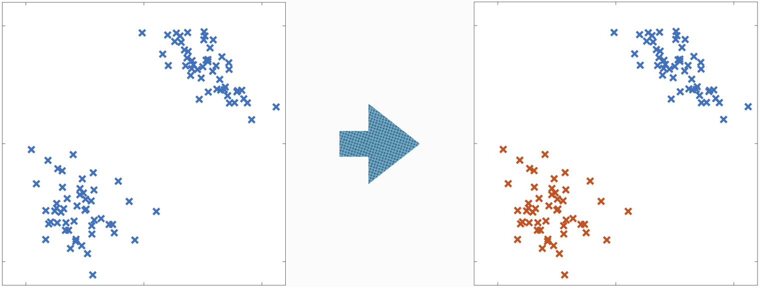
Datapoint Index	x_1	x_2	x_3	…	x_M	y
1	0.24	0.12	-0.34	…	0.98	3
2	0.56	1.22	0.20	…	1.03	2
3	-3.20	-0.01	0.21	…	0.93	1
…	…	…	…	…	…	…
N	2.24	1.76	-0.47	…	1.16	2

- Use this data to learn a predictive model $f_\theta : \mathcal{X} \to \mathcal{Y}$ (e.g. by optimizing θ)
- Once learned, we can use this to predict outputs for new input points, e.g. $f_\theta([0.48 \;\; 1.18 \;\; 0.34 \;\; \ldots \;\; 1.13]) \;=\; 2$

Unsupervised Learning

- In unsupervised learning we have no clear output variable that we are attempting to predict: $\mathcal{D} = \{x_i\}_{i=1}^{n}$
- This is sometimes referred to as **unlabeled data**
- Aim is to extract some salient features for the dataset, such as underlying structure, patterns, or characteristics
- Examples: clustering, feature extraction, density estimation, representation learning, data visualization, data compression

Unsupervised Learning—Clustering



Unlabeled Data

Group into Clusters

Unsupervised Learning—Deep Generative Models

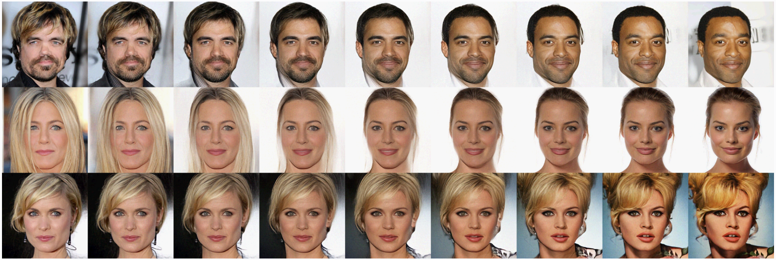
Learn powerful models for generating new datapoints

These are not real faces: they are samples from a learned model!

[1] D P Kingma and P Dhariwal. "Glow: Generative flow with invertible 1x1 convolutions". In: **NeurIPS**. 2018.

Unsupervised Learning—Deep Generative Models

Once learned, we can use this for a number of tasks like manipulating an image:



[1] Kingma and Dhariwal, "Glow: Generative flow with invertible 1x1 convolutions".

Other Types of Machine Learning

- Reinforcement learning
- Semi-supervised learning
- Meta-learning
- Online learning
- Active learning

Discriminative vs Generative Machine Learning

Discriminative vs Generative Machine Learning

- Discriminative methods **directly make predictions**
- Generative methods try to explain **how** the data was generated

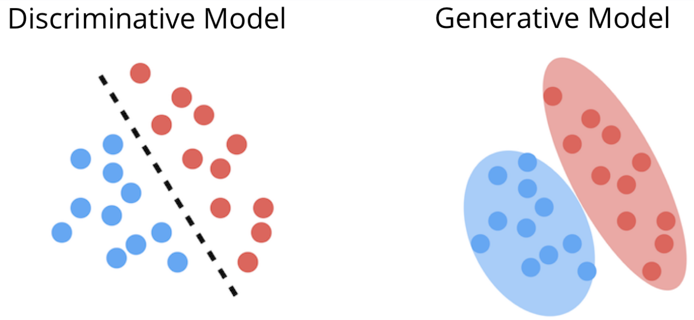


Image credit: Jason Martuscello, medium.com

Discriminative Machine Learning

- In supervised settings, discriminative methods **directly** learn a mapping f from inputs x to outputs y (or some characterization of y, e.g. class probabilities)
  - Though less common, there are some discriminative unsupervised approaches
- **Training** uses $\mathcal{D} = \{x_i, y_i\}_{i=1}^{n}$ to estimate the optimal predictive function f^* by minimizing an expected loss (more on this later)
- **Prediction** at a new input x involves simply applying the learned predictive function

Discriminative Machine Learning

Common approaches: neural networks, support vector machines, random forests, linear/logistic regression, k–nearest neighbors

Pros[2]

- Simpler to directly solve prediction problem than model the whole data generation process
- Typically make few assumptions
  - Often very effective for large datasets (e.g. deep learning, non–parametric approaches)
  - Some methods can be used effectively in a black-box manner

Cons

- Can be difficult to impart domain expertise
- Typically lack interpretability

[2] Note that all these pros and cons are generalizations and may not always hold

Generative Machine Learning

- Generative approaches construct a **probabilistic model** to explain **how** the data is generated
- For example, with labeled data $\mathcal{D} = \{x_i, y_i\}_{i=1}^n$, we might construct a joint model $p(X, Y)$ over inputs and outputs
- This in turns implies a predictive model via the conditional distribution $p(Y = y|X = x)$
- Can also be generative about the **model parameters** θ: e.g. with unsupervised data $\mathcal{D} = \{x_i\}_{i=1}^n$, we can construct a generative model $p(\theta, X)$
    - This is the foundation for Bayesian machine learning

Common approaches: mixture models, probabilistic context-free grammars, deep generative models, Boltzmann machines, Bayesian approaches

Pros[3]

- Allow us to make stronger modeling assumptions and thus incorporate more problem–specific expertise
- Provide explanation for how data was generated
    - More interpretable
    - Can provide additional information other than just prediction
- Many methods naturally provide uncertainty estimates

[3] Note the all these pros and cons are generalizations and may not always hold

Generative Machine Learning

Cons

- Tackling an inherently more difficult problem than straight prediction
- Can impart unwanted assumptions—often less effective for huge datasets
- Tend to require more problem–specific expertise

Recap

- Machine learning is all about **learning from data**
- Supervised learning has access to **outputs**, unsupervised learning does not
- Discriminative methods try and **directly** make predictions, generative methods try to explain **how** the data is generated

Further Reading

- Look at the course notes!
- Chapter 1 of K P Murphy. **Machine learning: a probabilistic perspective**. 2012.

  https://www.cs.ubc.ca/~murphyk/MLbook/pml-intro-22may12.pdf.
- L Breiman. "Statistical modeling: The two cultures". In: **Statistical science** (2001)