Probabilistic Machine Learning
— Course Recap —

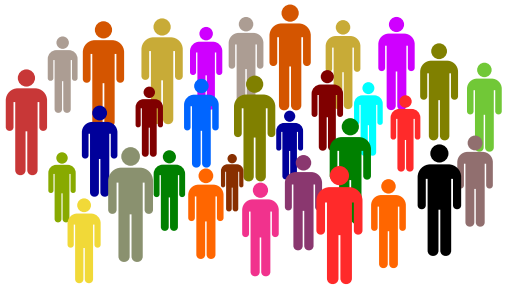
Michael U. Gutmann

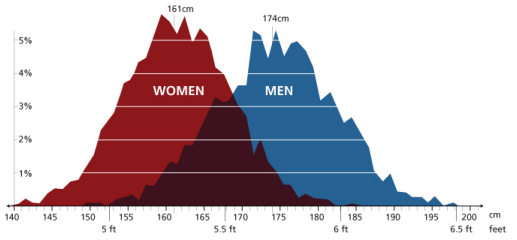
Probabilistic Machine Learning (INFR11298)
School of Informatics, The University of Edinburgh

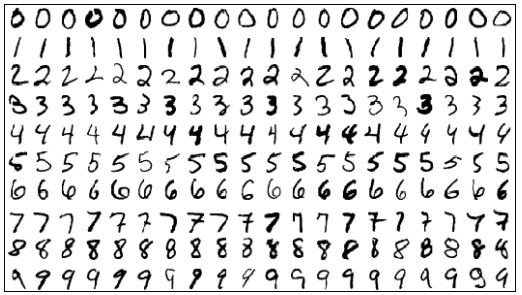
Autumn Semester 2026

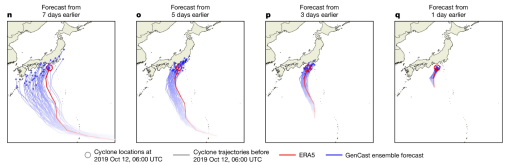
The beginning: variability and uncertainty

- We started the course with the observation that variability is part of nature.
- Variability leads to uncertainty when analysing or drawing conclusions from data: is it a 1 or a 7?
- Modelling and quantifying uncertainty is important in many domains, be it image generation or weather forecasting.









Probabilistic machine learning

- Probabilistic machine learning embraces uncertainty.
- We used probabilities to describe our beliefs of how things usually behave and how much they vary.
  → Topic 1: Representation
- We computed probabilities for things that we have not observed or cannot observe from things that we can observe.
  → Topic 2: Exact inference
- We combined beliefs with costs/utilities to choose actions that take uncertainty into account.
  → Topic 3: Actions and decision making
- We learned/updated beliefs from data.
  → Topic 4: Learning
- We saw that tracking all eventualities is often impossible.
  → Topic 5: Approximate inference and learning

The roadmap we followed

The probabilistic approach provides a principled framework, but direct implementation quickly runs into computational issues. Consider, for example,

$$p(\mathbf{x}|\mathbf{y}_o) = \frac{\sum_{\mathbf{z}} p(\mathbf{x}, \mathbf{y}_o, \mathbf{z})}{\sum_{\mathbf{x},\mathbf{z}} p(\mathbf{x}, \mathbf{y}_o, \mathbf{z})}$$

where $\mathbf{x}, \mathbf{y}, \mathbf{z}$ are each $d = 500$ dimensional, with each element taking $K = 10$ values.

- Issue 1: To specify $p(\mathbf{x}, \mathbf{y}, \mathbf{z})$, we need to specify $K^{3d} - 1 = 10^{1500} - 1$ non-negative numbers, which is impossible.

  Topic 1: Representation What reasonably weak assumptions can we make to efficiently represent $p(\mathbf{x}, \mathbf{y}, \mathbf{z})$?

The roadmap we followed

$$p(\mathbf{x}|\mathbf{y}_o) = \frac{\sum_{\mathbf{z}} p(\mathbf{x},\mathbf{y}_o,\mathbf{z})}{\sum_{\mathbf{x},\mathbf{z}} p(\mathbf{x},\mathbf{y}_o,\mathbf{z})}$$

- Issue 2: The sum in the numerator goes over the order of $K^d = 10^{500}$ non-negative numbers and the sum in the denominator over the order of $K^{2d} = 10^{1000}$, which is impossible to compute.

  Topic 2: Exact inference Can we further exploit the assumptions on $p(\mathbf{x}, \mathbf{y}, \mathbf{z})$ to efficiently compute the posterior probability or derived quantities?
- Issue 3: Thank you for the numbers. But what shall I best do?

  Topic 3: Actions and decision making How to predict the outcome of actions and choose optimal actions?

The roadmap we followed

- Issue 4: Where do the non-negative numbers $p(\mathbf{x}, \mathbf{y}, \mathbf{z})$ come from?

  Topic 4: Learning How can we learn the numbers from data?

- Issue 5: For some models, exact inference and learning is too costly even after fully exploiting the assumptions made.

  Topic 5: Approximate inference and learning

Topic 1: Representation

We discussed reasonably weak assumptions to efficiently represent the joint pdf/pmf $p(\mathbf{x}, \mathbf{y}, \mathbf{z})$ of many variables.

- Two classes of assumptions: independence and parametric assumptions
- They are orthogonal to each other and can be combined
- Directed and undirected graphical models
- Independencies encoded by the graphs and their expressive power
- Connection between independencies and the factorisation of the pdf/pmf
- The chain rule and autoregressive models

Topic 2: Exact inference

We have seen that the independence assumptions allow us, under certain conditions, to efficiently compute conditional/posterior distributions or derived quantities.

- Factor graphs to represent factorisations
- Variable elimination for general factor graphs
- Inference when the model can be represented as a factor tree (message passing algorithms)
- Application to Hidden Markov models

Topic 3: Actions and decision making

We then discussed how to turn (conditional) distributions into actions.

- We modelled actions as interventions in directed graphical models and used decision theory to derive optimal actions
- Causal DAGs and how interventions change their structure
- Methods to compute the effect of interventions, i.e. the postinterventional distribution
- Turning beliefs into actions by minimising the expected loss
- Standard loss functions and the corresponding optimal actions
- Learning as decision making under uncertainty

Topic 4: Learning

We discussed methods to learn probabilistic models from data.

- Learning by Bayesian inference
- Learning by parameter estimation
- Likelihood function
- Factor analysis and independent component analysis

Topic 5: Approximate inference and learning

We discussed computational intractabilities in inference and learning: intractable integrals often preclude exact inference and likelihood-based learning.

- Intractable integrals may be due to unobserved variables or intractable partition functions
- Variational approaches to learning and inference, including mean-field VI and the EM algorithm
- Application of the EM algorithm to Hidden Markov Models
- Learning of deep latent variable models and variational autoencoders
- Monte Carlo integration and sampling