Introduction to Machine Learning

Mahesan Niranjan
University of Southampton

DiscNet Summer School

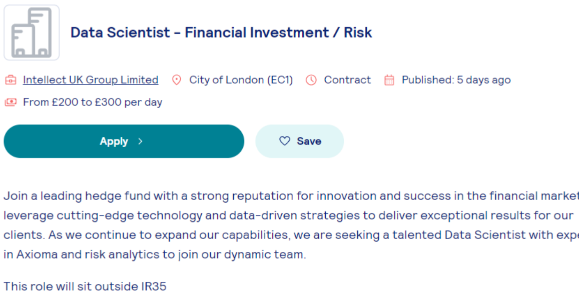
June 2025

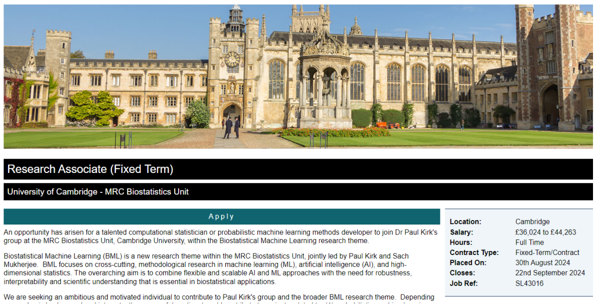
Modules We Offer in Southampton

- Foundations of Machine Learning
- Machine Learning Technologies
- Advanced Machine Learning
- Differential Programming and Deep Learning
- Deep Learning Technologies
- Bayesian, Active & Reinforcement Learning
- Natural Language Processing
- Causal Reasoning and Machine Learning
- Computational Biology
- Computer Vision
- Computational Finance

- This week is a random sample from magenta

Machine Learning: Good employment prospects!





Data scientist in quantitative finance
Intellect UK Group Ltd.
Hedge Fund
150 - 200K

Researcher in Biostatistics
Medical Research Council, Cambridge
Cancer Research
36 - 44K

`Standard disclaimers apply!`

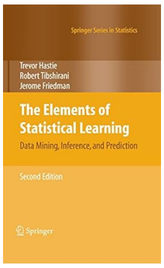
Foundations of Machine Learning
… draws from several disciplines

- Function approximation
- Linear algebra
  - Low rank approximation, Singular Value Decomposition
- Probability and Statistics
  - Multi-variate Gaussian, Information Theory, Bayes' Formula
- Optimization
- Software Engineering
- Parallel Computing
- Simulation, Experiment Design

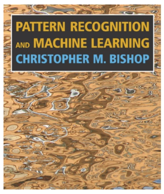
Machine Learning: Data-driven Modelling

Data	$\{\boldsymbol{x}_n,\ \boldsymbol{t}_n\}_{n=1}^{N} \qquad \{\boldsymbol{x}_n\}_{n=1}^{N}$
Function Approximator	$\boldsymbol{t} \ = \ f\,(\boldsymbol{x},\ \boldsymbol{\theta}) \ + \ v$
Parameter Estimation	$E_0 \ = \ \sum_{n=1}^{N} \{\|\|\ \boldsymbol{t}_n \ - \ f\,(\boldsymbol{x}_n;\ \boldsymbol{\theta})\ \|\|\}^2$
Prediction	$\hat{\boldsymbol{t}}_{N+1} \ = \ f\left(\boldsymbol{x}_{N+1},\ \hat{\boldsymbol{\theta}}\right)$
Regularization	$E_1 \ = \ \sum_{n=1}^{N} \{\|\|\ \boldsymbol{t}_n \ - \ f\,(\boldsymbol{x}_n)\ \|\|\}^2 \ + \ r\,(\|\|\boldsymbol{\theta}\|\|)$
Modelling Uncertainty	$p\left(\boldsymbol{\theta} \mid \{\boldsymbol{x}_n, \boldsymbol{t}_n\}_{n=1}^{N}\right)$
Probabilistic Inference	$\boldsymbol{E}\,[g\,(\boldsymbol{\theta})] = \int g\,(\boldsymbol{\theta})\,p\,(\boldsymbol{\theta})\,d\boldsymbol{\theta} \ = \frac{1}{N_s}\sum_{n=1}^{N_s} g\left(\boldsymbol{\theta}^{(n)}\right)$
Sequential Estimation	$\boldsymbol{\theta}\,(n-1 \mid n-1) \longrightarrow \boldsymbol{\theta}\,(n \mid n-1) \longrightarrow \boldsymbol{\theta}\,(n \mid n)$ Kalman & Particle Filters; Reinforcement Learning

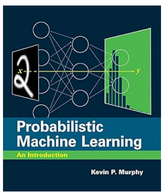
Good Books
To study the fundamentals



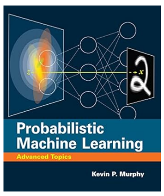
T. Hastie, R. Tibshirani & J. Friedman
Elements of Statistical Learning



C.M. Bishop
PRML : Pattern Recognition and Machine Learning



K. Murphy
Proabilistic Machine Learning



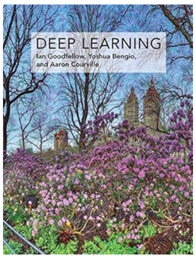
Advanced Topics

> *"There is nothing to be learnt from a professor, which is not to be met with in books"*
> *- David Hume (1711-1776)*

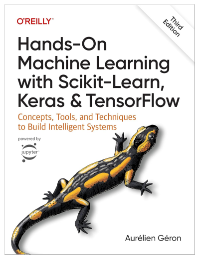
(**WikiPedia:** "Hume had little respect for the professors of his time [...] He did not graduate")

More books...

For practical aspects of the subject...

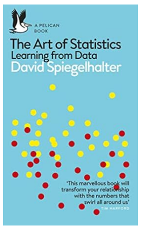


I. Goodfellow *et al.*
Deep Learning

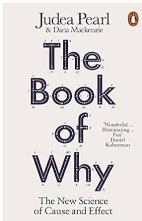


A. Geron
Handson Machine Learning with Scikit-Learn

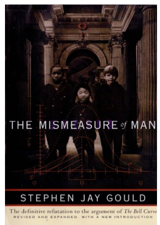
For some general, societally important, aspects of inference from data...



D. Spiegelhalter
The Art of Statistics



J. Pearl
The Book of Why



S.J. Gould
Mismeasure of Man

Machine Learning: Data-driven Modlling

Data	$\{\boldsymbol{x}_n,\ \boldsymbol{t}_n\}_{n=1}^N \qquad \{\boldsymbol{x}_n\}_{n=1}^N$
Function Approximator	$\boldsymbol{t} \ = \ f\left(\boldsymbol{x},\ \boldsymbol{\theta}\right) \ + \ v$
Parameter Estimation	$E_0 \ = \ \sum_{n=1}^N \left\{\|\| \ \boldsymbol{t}_n \ - \ f\left(\boldsymbol{x}_n;\ \boldsymbol{\theta}\right) \ \|\|\right\}^2$
Prediction	$\hat{\boldsymbol{t}}_{N+1} \ = \ f\left(\boldsymbol{x}_{N+1},\ \hat{\boldsymbol{\theta}}\right)$
Regularization	$E_1 \ = \ \sum_{n=1}^N \left\{\|\| \ \boldsymbol{t}_n \ - \ f\left(\boldsymbol{x}_n\right) \ \|\|\right\}^2 \ + \ r\left(\|\|\boldsymbol{\theta}\|\|\right)$
Modelling Uncertainty	$p\left(\boldsymbol{\theta} \middle\| \left\{\boldsymbol{x}_n, \boldsymbol{t}_n\right\}_{n=1}^N\right)$
Probabilistic Inference	$\boldsymbol{E}\left[g\left(\boldsymbol{\theta}\right)\right] = \int g\left(\boldsymbol{\theta}\right) p\left(\boldsymbol{\theta}\right) d\boldsymbol{\theta} \ = \ \frac{1}{N_s} \sum_{n=1}^{N_s} g\left(\boldsymbol{\theta}^{(n)}\right)$
Sequential Estimation	$\boldsymbol{\theta}\left(n-1 \| n-1\right) \longrightarrow \boldsymbol{\theta}\left(n \| n-1\right) \longrightarrow \boldsymbol{\theta}\left(n \| n\right)$ Kalman & Particle Filters; Reinforcement Learning

Machine Learning
Some Recent Challenges, Trends

- Function approximator: Shallow versus deep neural architectures
- Several tricks of parameter estimation of large (and deep) models
- Several tricks to induce generalization
- Automatic differentiation
- Problem formulation
  - Supervised & unsupervised versus Self-supervised learning
  - Human-in-the-loop Reinforcement Learning
  - Combining Mechanistic and Data-driven models – *e.g.* Physics Informed Neural Networks
- Interesting problems in various domains.
- Large amounts of data (sometimes illegally acquired!)
- Large amounts of computing (only a few can afford!)

Lab One: Multi-variate Gaussian

- Univariate Gaussian

$$p(x) \;=\; \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{(x-m)^2}{\sigma^2}\right\}$$

- Multivariate Gaussian $\boldsymbol{x} \in \mathcal{R}^D$

$$p\left(\boldsymbol{x}\right) \;=\; \frac{1}{(2\pi)^{D/2}(\det \boldsymbol{\Sigma})^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{m})^t\,\boldsymbol{\Sigma}^{-1}\left(\boldsymbol{x}-\boldsymbol{m}\right)\right\}$$

  Mean $\boldsymbol{m}$ is a vector
  Covariance, C, matrix: symmetric, positive definite!

$$\boxed{\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{m}, \boldsymbol{\Sigma}),\; \boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} \;\Longrightarrow\; \boldsymbol{y} \sim \mathcal{N}(\boldsymbol{A}\boldsymbol{m},\, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T)}$$

- Sampling from a given density
- Projections of data onto a subspace