Advanced Machine Learning
Lecture 1: Introduction

Sandjai Bhulai
Vrije Universiteit Amsterdam

s.bhulai@vu.nl
6 September 2023

VU VRIJE UNIVERSITEIT AMSTERDAM
Faculty of Science

Motivation

Metronieuws

10.000 mensen bij protest op de Dam, gemeente vraagt niet te komen

1 dag geleden

Het Parool

Grote drukte bij protest op de Dam, gemeente roept op niet te komen

2 dagen geleden

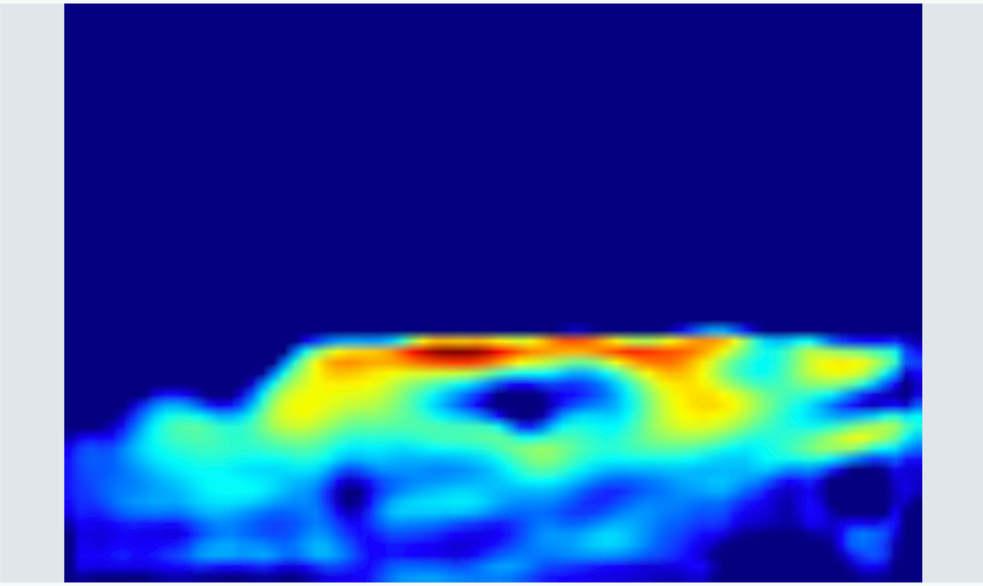
Het Parool

20.000 coronabetogers op de Dam: 'Ik voel me vreemd, niet veilig en buitengesloten'

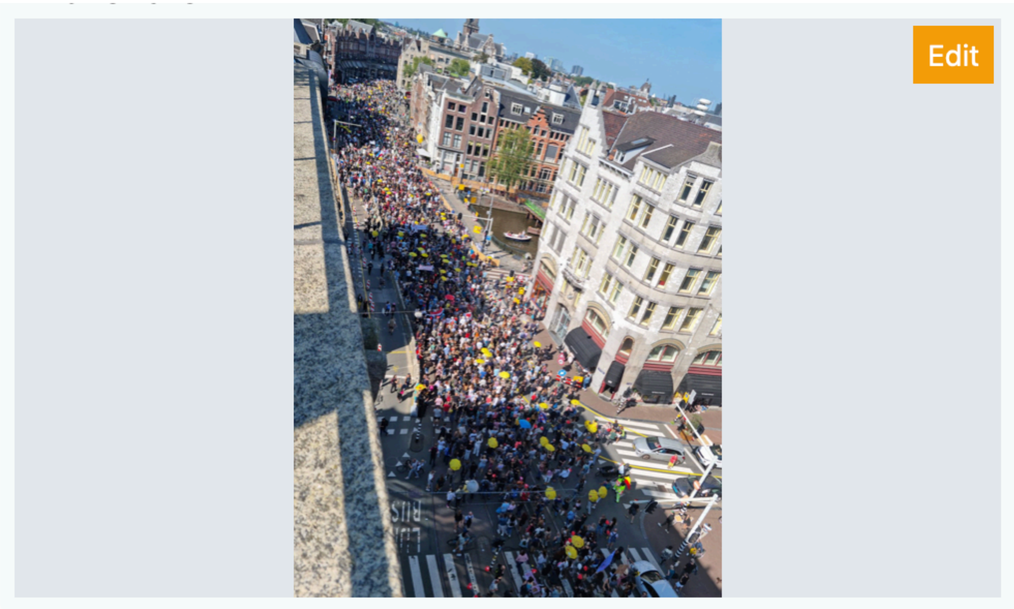
1 dag geleden

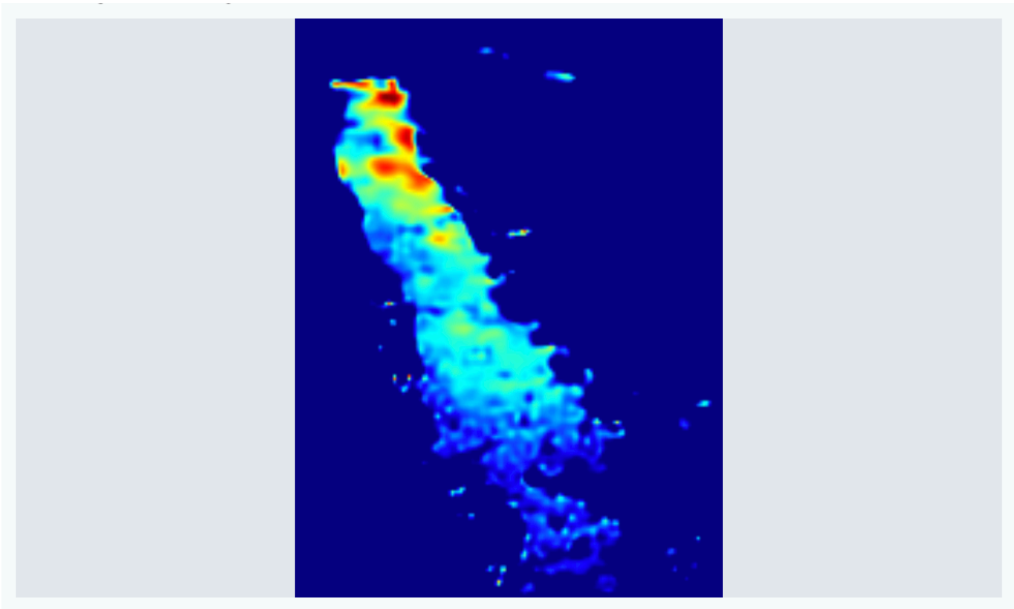
→ Alles bekijken

Motivation



Motivation





Motivation

a professor teaching the course advanced machine learning

Motivation

S what skills does an machine learning professional need to have?

The skill set required for a machine learning professional can be quite extensive and may vary depending on the specific role, industry, and type of projects one is working on. However, some fundamental skills are almost universally required.

Technical Skills:

1. **Programming Languages**:
   - Proficiency in Python is often essential, given its widespread use in the field of machine learning.
   - Knowledge of languages like R, Java, or C++ can also be beneficial for specific tasks or environments.
2. **Machine Learning Frameworks**:
   - Familiarity with libraries and frameworks like TensorFlow, PyTorch, Keras, and Scikit-learn is important for implementing algorithms and models.
3. **Data Manipulation and Analysis**:
   - Skills in data preprocessing and transformation using libraries like Pandas, NumPy, and data visualization libraries like Matplotlib, Seaborn, or Plotly.
4. **Statistics and Mathematics**:
   - Strong understanding of statistics, probability, linear algebra, calculus, and discrete mathematics.

Motivation

5. **Data Wrangling**:
   - Ability to handle, clean, and organize data, sometimes referred to as data munging or data wrangling.
6. **Feature Engineering**:
   - The ability to select the most appropriate features for training models or create new features from existing ones.
7. **Model Evaluation**:
   - Familiarity with metrics to evaluate machine learning models like precision, recall, F1 score, ROC AUC, etc.
8. **Natural Language Processing**:
   - Text processing skills may be required if working in areas that involve understanding human language.
9. **Computer Vision**:
   - Familiarity with image processing techniques and libraries like OpenCV can be essential for some roles.
10. **Optimization Techniques**:
    - Understanding of optimization algorithms used in machine learning, such as stochastic gradient descent.
11. **Big Data Technologies**:
    - Familiarity with big data technologies like Hadoop and Spark can be beneficial when dealing with large datasets.
12. **Cloud Services**:
    - Know-how to deploy machine learning models using cloud services like AWS, Google Cloud, or Azure.

Machine learning

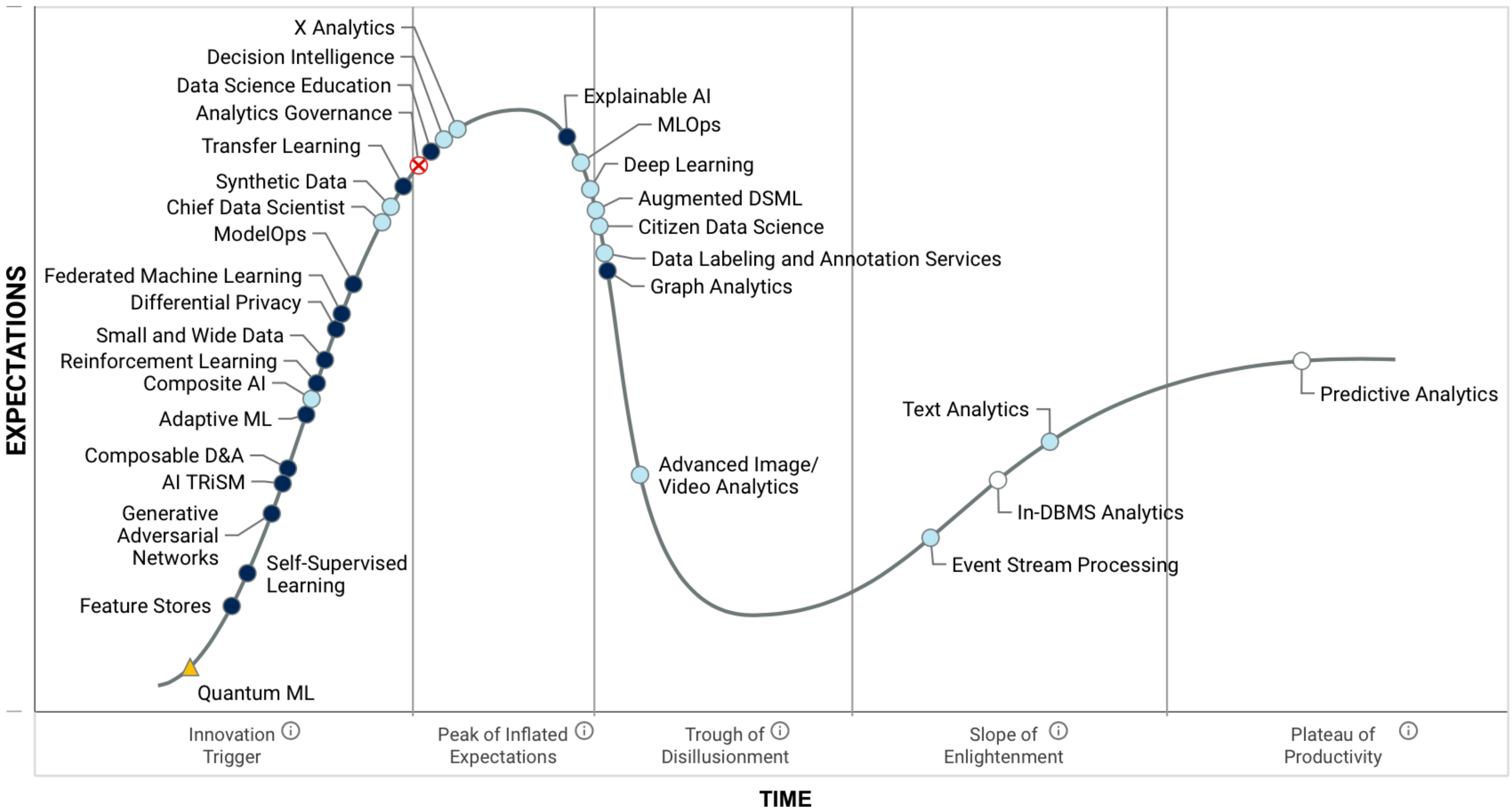
Machine learning

What is machine learning?

- Arthur Samuel (1959): field of study that gives computers the ability to learn without being explicitly programmed

- Tom Mitchell (1998): a computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E

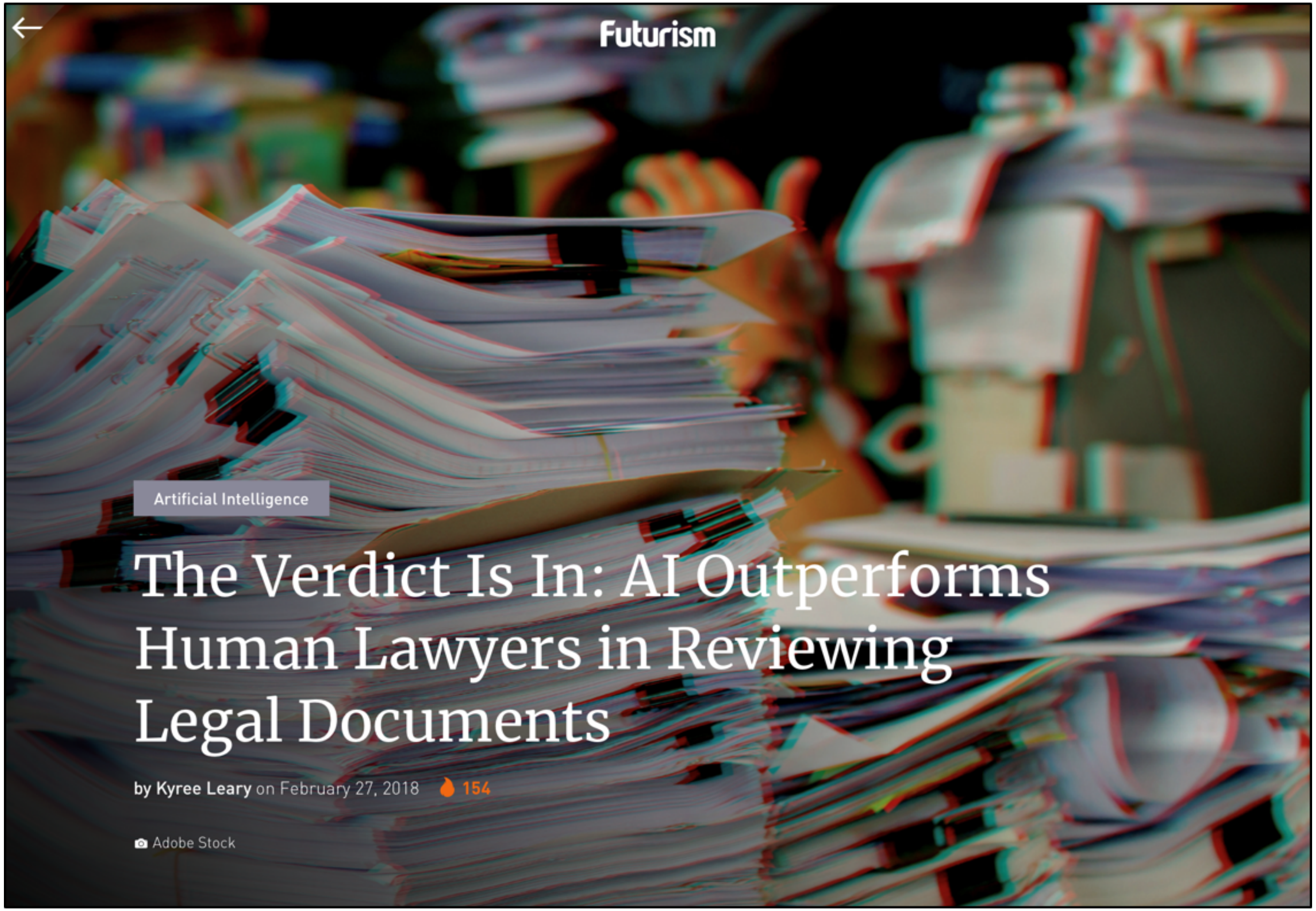
Machine learning



Machine learning examples

- Database mining
  - Web click data, medical records, biology

Machine learning examples



Machine learning examples

- Database mining
  - Web click data, medical records, biology

- Applications that cannot be programmed by hand
  - Autonomous vehicles, OCR, NLP, computer vision

Machine learning examples

Google's new AI algorithm predicts heart disease by looking at your eyes

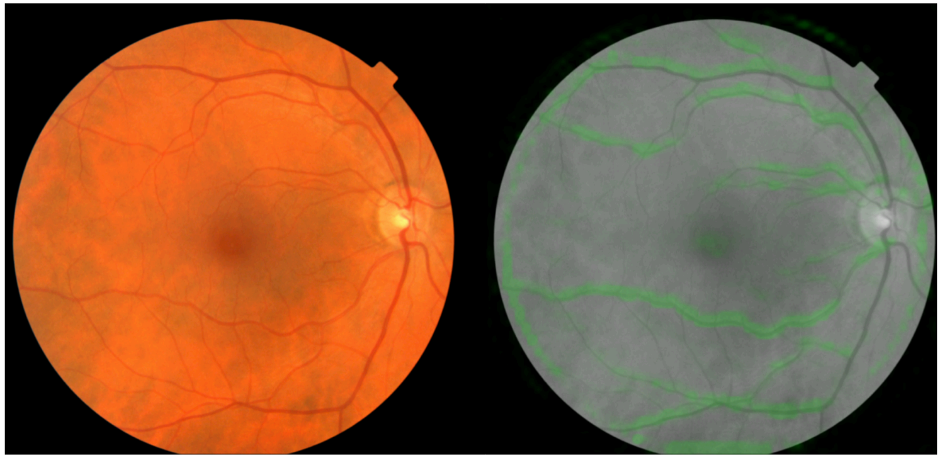
9

Experts say it could provide a simpler way to predict cardiovascular risk

By James Vincent | @jjvincent | Feb 19, 2018, 12:04pm EST

SHARE

The algorithm could allow doctors to predict cardiovascular risk more simply by using scans of the retina. | Stock photo by Scott Olson/Getty Images

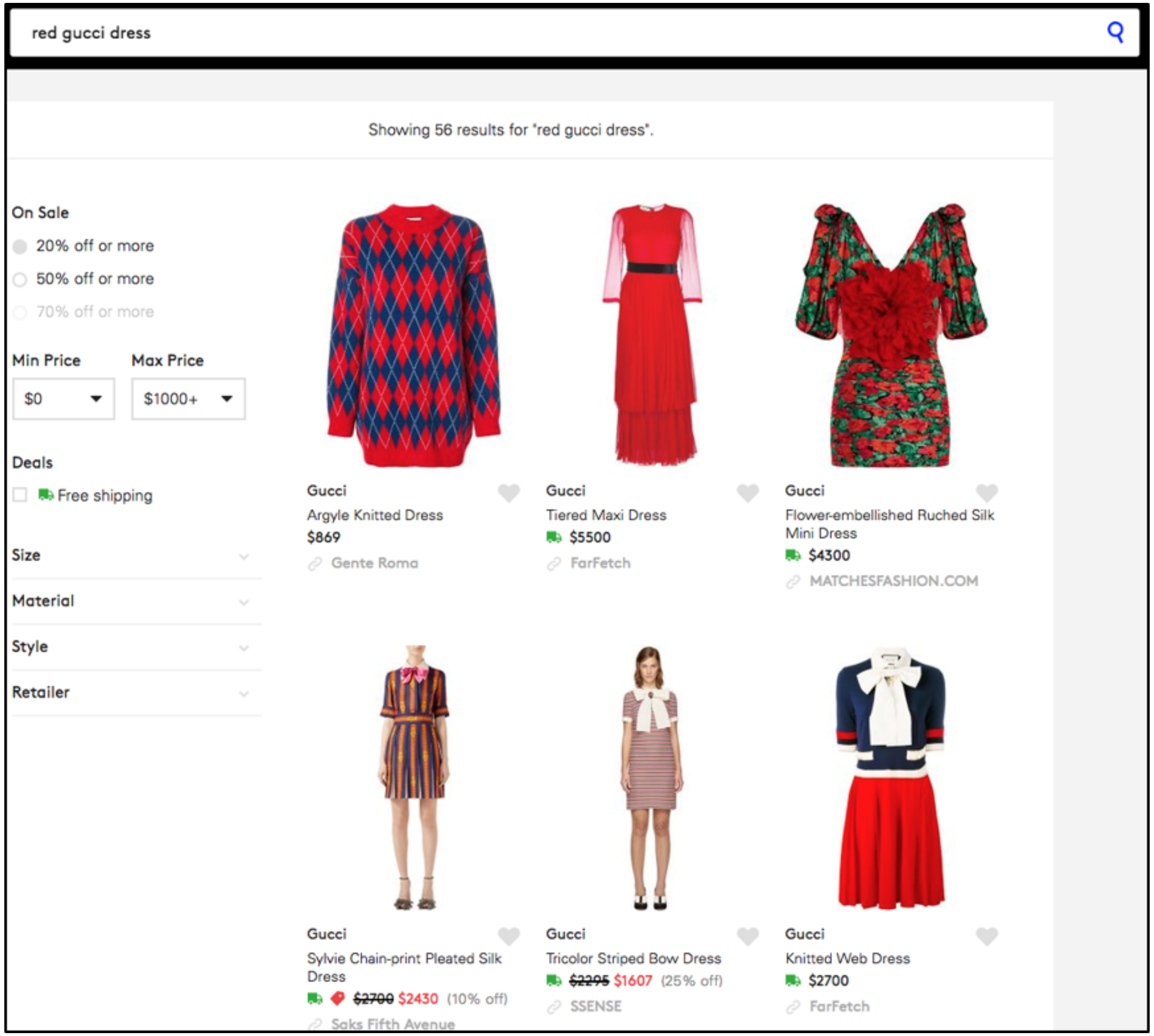


Two images of the fundus, or interior rear of your eye. The one on the left is a regular image; the on the right shows how Google's algorithm picks out blood vessels (in green) to predict blood pressure. | Photo by Google / Verily Life Sciences

Machine learning examples

- Database mining
  - Web click data, medical records, biology
- Applications that cannot be programmed by hand
  - Autonomous vehicles, OCR, NLP, computer vision
- Self-customizing programs
  - Amazon, Netflix product recommendations

Machine learning examples



Machine learning examples

- Database mining
  - Web click data, medical records, biology
- Applications that cannot be programmed by hand
  - Autonomous vehicles, OCR, NLP, computer vision
- Self-customizing programs
  - Amazon, Netflix product recommendations
- Understanding human learning / behavior
  - Brain, real AI

Machine learning examples

Is Apple the Next Big Life Insurance Company?

Course overview

Advanced Machine Learning

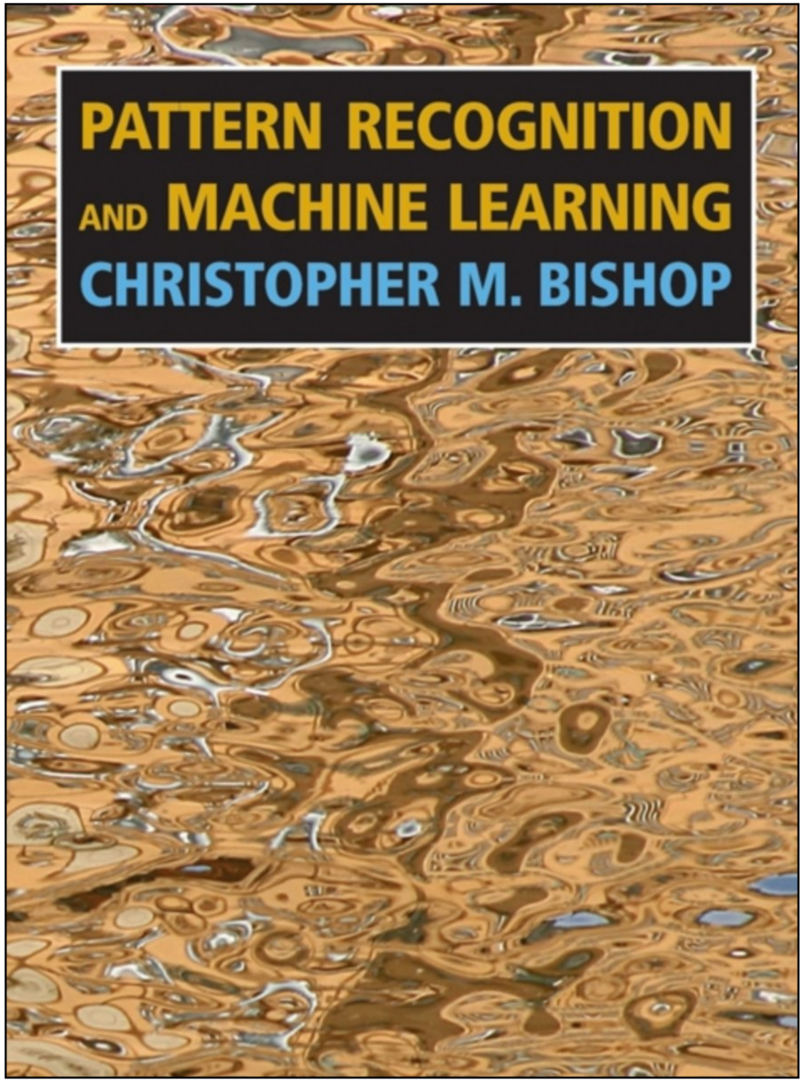
Course overview: objectives

- Understand the capabilities and the limitations of machine learning
- Implement machine learning algorithms in Python
- Know relevant machine learning algorithms for both supervised and unsupervised learning problems
- Select the right machine learning models for real-world use cases
- Understand when to apply online learning, reinforcement learning, and deep learning
- Interpret the outcomes of machine learning algorithms

Course overview: grading

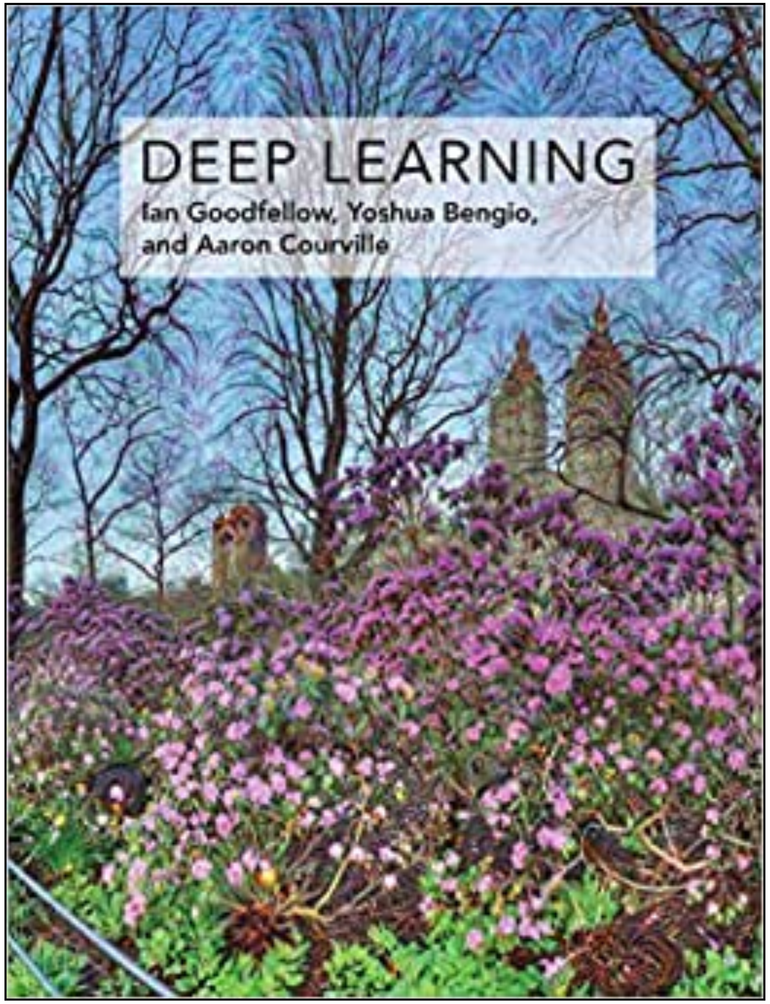
- Course value: 6 ECTS (a full month of work)
- Course consists of:
  - Written exam (weight: 0.9)
  - Computer assignments (weight: 0.1)
- Both parts have to be passed with at least a 5.5
- Written assignments for exam preparation
- Q&A sessions on Wednesday for written assignments
- Teaching assistant: Mathijs Pellemans
- Slack workspace

Course overview: books



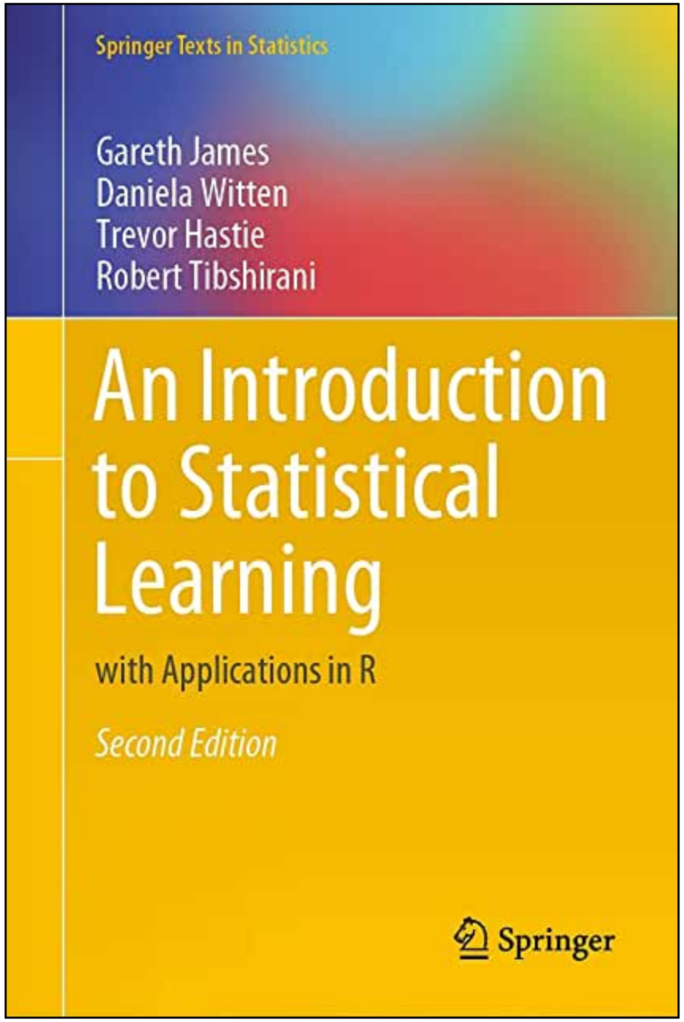
- Book presents mathematical theory behind machine learning
- *Very* mathematical
- Lectures will try to explain the mathematics

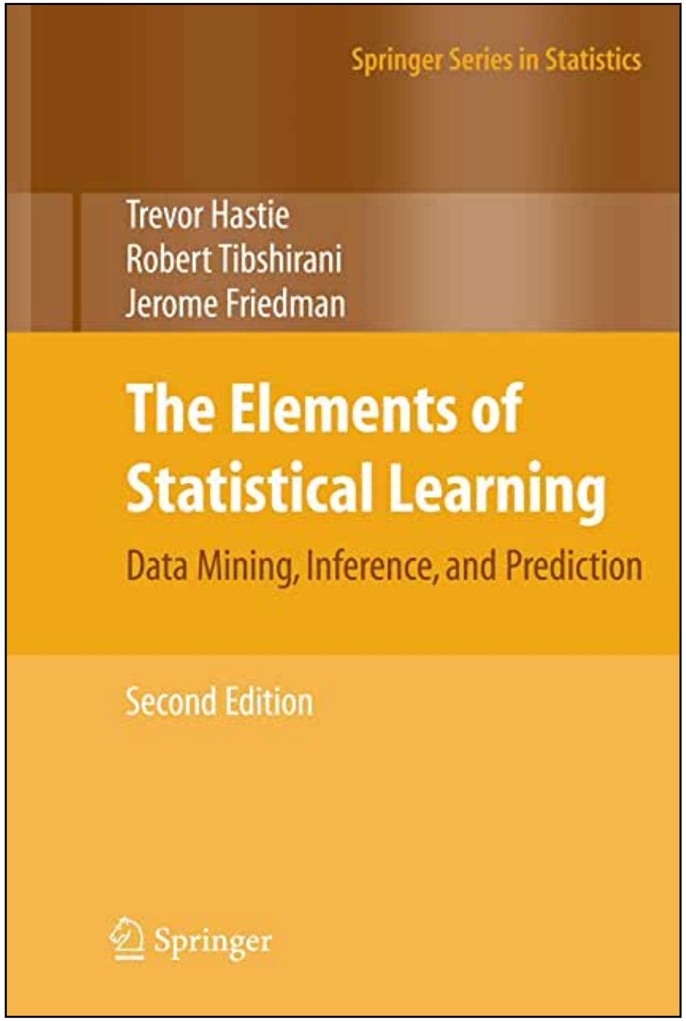
Course overview: books



- Book presents concepts behind deep learning
- *Not* mathematical
- Lectures will try to explain the mathematics

Course overview: books





Course overview: program

- Regression models (linear, logistic, etc.)
- Neural networks (NNs)
- Convolutional NNs
- Recurrent NNs
- Hidden Markov models
- Reinforcement learning

Course overview: software

ANACONDA

jupyter

python™

Keras + TensorFlow

Preliminaries

Advanced Machine Learning

Linear algebra

- A matrix A has elements A_{ij} where i indexes rows and j indexes columns

Properties of matrices:

- $AA^{-1} = A^{-1}A = I$
- $(AB)^\top = B^\top A^\top$
- $(AB)^{-1} = B^{-1}A^{-1}$
- $(A^\top)^{-1} = (A^{-1})^\top$
- $(A^\top)^\top = A$
- $(A + B)^\top = A^\top + B^\top$

Linear algebra

- Symmetric matrix: $A^\top = A$
- Positive definite: $x^\top A x > 0$ for all non-zero x and A symmetric

- All vectors in this course will be column vectors
- Recall, if $A \in \mathbb{R}^{m \times n}$ is a matrix and $\mathbf{x} \in \mathbb{R}^n$ a vector, then the i-th element of $A\mathbf{x}$ is given by $(A\mathbf{x})_i = \sum_{j=1}^{n} a_{ij} x_j$

Vector calculus

- Let $\mathbf{a} \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}^n$ be a vector, and $x \in \mathbb{R}$ a scalar. Then

$$\left(\frac{\partial \mathbf{a}}{\partial x}\right)_i = \frac{\partial a_i}{\partial x}$$

Derivatives with respect to a vector are defined by

- scalar-vector $$\left(\frac{\partial x}{\partial \mathbf{a}}\right)_i = \frac{\partial x}{\partial a_i}$$
- vector-vector $$\left(\frac{\partial \mathbf{a}}{\partial \mathbf{b}}\right)_{ij} = \frac{\partial a_i}{\partial b_j}$$

Vector calculus

- Example: let $\mathbf{a} \in \mathbb{R}^n$ and $x \in \mathbb{R}^n$. What is $\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^\top \mathbf{a})$?

- Solution: this is a scalar-vector derivative, thus

$$\left(\frac{\partial \mathbf{x}^\top \mathbf{a}}{\partial \mathbf{x}}\right)_i = \frac{\partial \mathbf{x}^\top \mathbf{a}}{\partial x_i} = \frac{\partial \sum_{j=1}^n a_j x_j}{\partial x_i} = a_i$$

- Thus $\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^\top \mathbf{a}) = \mathbf{a}^\top$, similarly $\frac{\partial}{\partial \mathbf{x}}(\mathbf{a}^\top \mathbf{x}) = \mathbf{a}^\top$

Vector calculus

- Another example. Let $\mathbf{b} \in \mathbb{R}^m$, $A \in \mathbb{R}^{m \times n}$, and $\mathbf{x} \in \mathbb{R}^n$
- What is $\dfrac{\partial \mathbf{b}^\top A\mathbf{x}}{\partial \mathbf{x}}$?
- Solution: this is a scalar-vector derivative, thus

$$\left(\frac{\partial \mathbf{b}^\top A\mathbf{x}}{\partial \mathbf{x}}\right)_i = \frac{\partial \mathbf{b}^\top A\mathbf{x}}{\partial x_i} = \frac{\partial \sum_{k=1}^{n} \sum_{j=1}^{m} b_j a_{jk} x_k}{\partial x_i} = \sum_{j=1}^{m} b_j a_{ji} = \sum_{j=1}^{m} a_{ij}^\top b_j = (A^\top \mathbf{b})_i$$

- Hence, $\dfrac{\partial \mathbf{b}^\top A\mathbf{x}}{\partial \mathbf{x}} = \mathbf{b}^\top A$

Vector calculus

- $\dfrac{\partial \mathbf{x}}{\partial \mathbf{x}} = I$
- $\dfrac{\partial A\mathbf{x}}{\partial \mathbf{x}} = A$
- $\dfrac{\partial \mathbf{x}^\top A}{\partial \mathbf{x}} = A^\top$
- $\dfrac{\partial \mathbf{x}^\top A\mathbf{x}}{\partial \mathbf{x}} = \mathbf{x}^\top(A + A^\top)$ and $2\mathbf{x}^\top A$ if A is symmetric
- $\dfrac{\partial^2 \mathbf{x}^\top A\mathbf{x}}{\partial \mathbf{x}\partial \mathbf{x}^\top} = A + A^\top$ and $2A$ if A is symmetric

Vector calculus

- $\dfrac{\partial \mathbf{x}}{\partial \mathbf{x}} = I$

- Proof:

$$\left(\frac{\partial \mathbf{x}}{\partial \mathbf{x}}\right)_{ij} = \frac{\partial x_i}{\partial x_j} = \begin{cases} 0, & \text{if } i \neq j, \\ 1, & \text{if } i = j \end{cases}$$

Vector calculus

- $$\frac{\partial A\mathbf{x}}{\partial \mathbf{x}} = A$$

- Proof:

$$\left(\frac{\partial A\mathbf{x}}{\partial \mathbf{x}}\right)_{ij} = \frac{\partial (A\mathbf{x})_i}{\partial x_j} = \frac{\partial \sum_{k=1}^{n} a_{ik} x_k}{\partial x_j} = a_{ij}$$

Vector calculus

- $$\frac{\partial \mathbf{x}^\top A \mathbf{x}}{\partial \mathbf{x}} = \mathbf{x}^\top (A + A^\top)$$

- Proof:

$$\frac{\partial \mathbf{x}^\top A \mathbf{x}}{\partial \mathbf{x}_i} = \frac{\partial \sum_{k=1}^{n} \left(\sum_{l=1}^{n} a_{kl} x_l \right) x_k}{\partial x_i} = \sum_{k \neq i} a_{ki} x_k + \sum_{l \neq i} a_{il} x_l + 2 a_{ii} x_i =$$

$$\sum_{k=1}^{n} a_{ki} x_k + \sum_{l=1}^{n} a_{il} x_l = (A^\top \mathbf{x})_i + (A\mathbf{x})_i$$

Vector calculus

- If $\mathbf{a}, \mathbf{b}, \mathbf{x} \in \mathbb{R}^n$, then $\dfrac{\partial \mathbf{a}^\top \mathbf{x}\mathbf{x}^\top \mathbf{b}}{\partial \mathbf{x}} = \mathbf{x}^\top(\mathbf{a}\mathbf{b}^\top + \mathbf{b}\mathbf{a}^\top)$

- Proof:

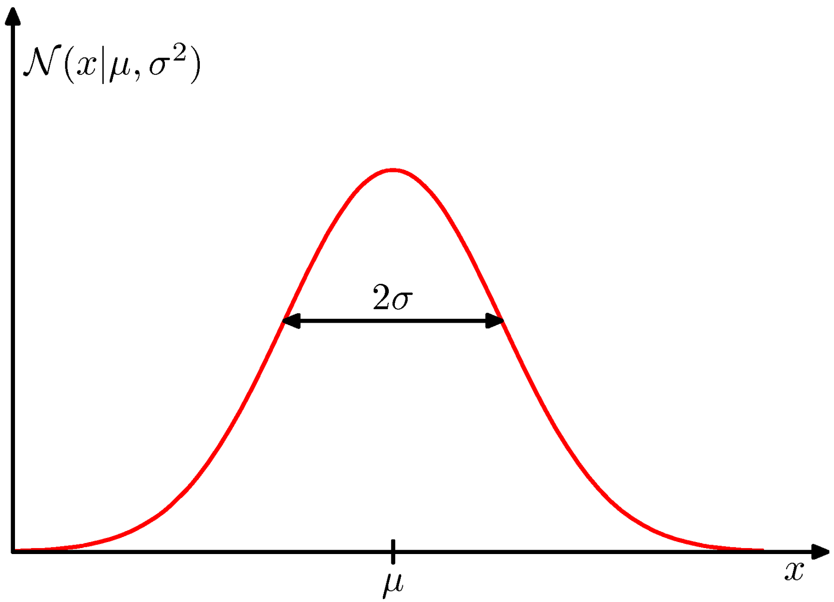
$$\frac{\partial \mathbf{a}^\top \mathbf{x}\mathbf{x}^\top \mathbf{b}}{\partial \mathbf{x}_i} = \frac{\partial \sum_{k=1}^n a_k x_k \sum_{l=1}^n x_l b_l}{\partial \mathbf{x}_i} = b_i \sum_{k=1}^n a_k x_k + a_i \sum_{l=1}^n x_l b_l =$$

$$b_i \mathbf{a}^\top \mathbf{x} + a_i \mathbf{x}^\top \mathbf{b} = b_i \mathbf{a}^\top \mathbf{x} + a_i \mathbf{b}^\top \mathbf{x} = \left(b_i \mathbf{a}^\top + a_i \mathbf{b}^\top\right) \mathbf{x}$$

Normal distribution

- Univariate Normal distribution

$$\mathcal{N}(x\,|\,\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$



$$\beta = 1/\sigma^2$$

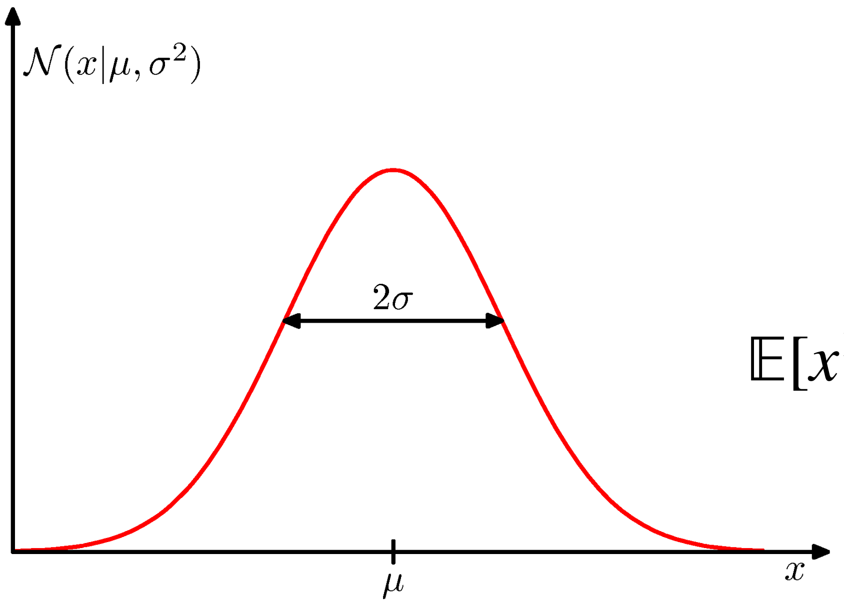
$$\mathcal{N}(x\,|\,\mu,\sigma^2) > 0$$

$$\int_{-\infty}^{\infty} \mathcal{N}(x\,|\,\mu,\sigma^2)\,\mathrm{d}x = 1$$

Normal distribution

- Univariate Normal distribution

$$\mathcal{N}(x\,|\,\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$

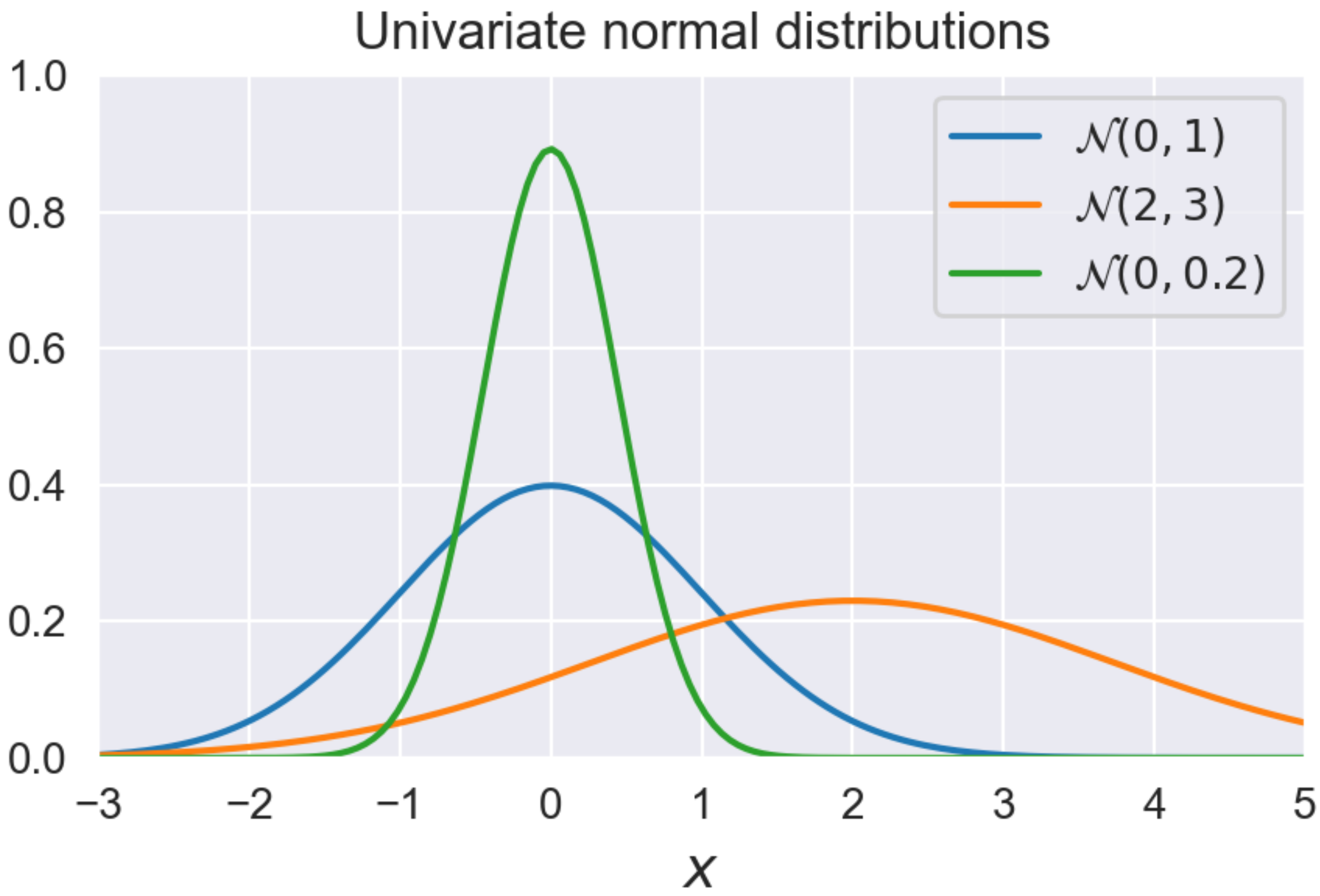


$$\mathbb{E}[x] = \int_{-\infty}^{\infty} x\mathcal{N}(x\,|\,\mu,\sigma^2)\,\mathrm{d}x = \mu$$

$$\mathbb{E}[x^2] = \int_{-\infty}^{\infty} x^2\mathcal{N}(x\,|\,\mu,\sigma^2)\,\mathrm{d}x = \mu^2 + \sigma^2$$

$$\mathrm{var}[x] = \mathbb{E}[x^2] - \mathbb{E}[x]^2 = \sigma^2$$

Normal distribution



Normal distribution

- Multivariate Normal distribution

$$\mathcal{N}(x \mid \mu, \Sigma) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right\}$$

- μ is a D-dimensional mean vector
- Σ is a $D \times D$ covariance matrix (symmetric, positive definite)
- $|\Sigma|$ is the determinant of Σ

Normal distribution

$$\mathcal{N}\left(\begin{bmatrix}0\\0\end{bmatrix}, \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}\right) \qquad \mathcal{N}\left(\begin{bmatrix}0\\1\end{bmatrix}, \begin{bmatrix}1 & 0.8\\0.8 & 1\end{bmatrix}\right)$$