Statistical Learning for Data Science 2

229352 · Semester 2/2026 · Donlapark Ponnoprat

Course information

- **Instructor:** Donlapark Ponnoprat
- **Lectures:** Friday, 9:00 to 11:00, SCB4202
- **Labs:** Tuesday, 9:00 to 11:00, SCB4405-06
- **Website:** donlapark.pages.dev/229352
  - schedule, labs and slides
- **Canvas:** instructure.com/courses/32929
  - announcements, labs and grades

Prerequisites

- **You should be comfortable with:**
  - linear algebra (matrix operations)
  - probability
  - programming, preferably in Python
- **Calculus:** recommended, but not required
- We use scikit-learn and PyTorch.

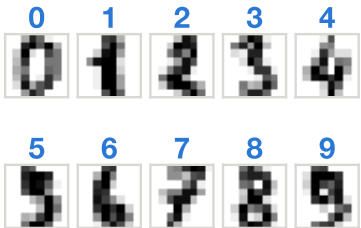
Grading

- **30%** labs (10 to 12 assignments)
- **35%** midterm exam
  - TBA. Please fill out the excel file in Canvas
- **35%** final exam
  - Friday, March 26, 12:00 to 15:00
- Your final grade may be curved.

Labs

- **Submit:** a link to colab or a notebook file (ipynb)
- **Deadline:** 23:59 on the next Monday
- Late labs' scores will be halved.
- Your lowest lab score is dropped.

What this course is about



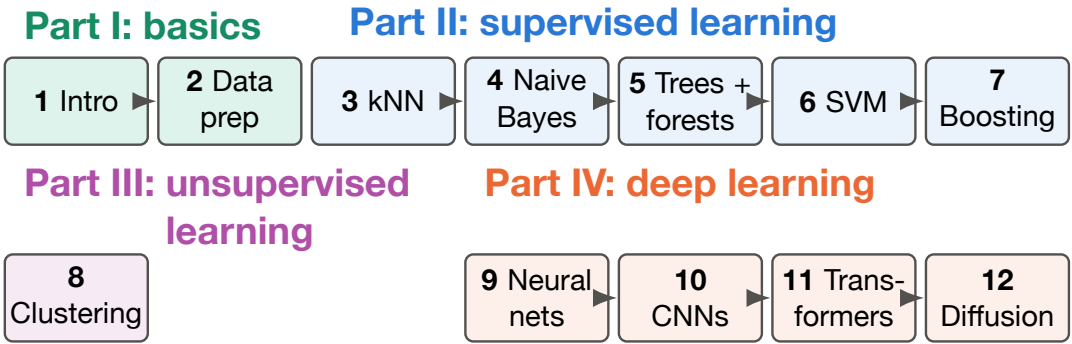
supervised learning
handwritten digits



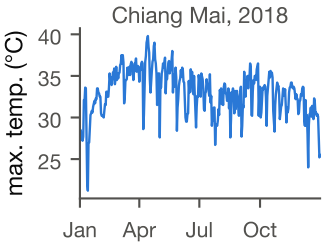
unsupervised learning
groups of similar digits

Digits: scikit-learn `load_digits` (UCI, E. Alpaydin and C. Kaynak), CC BY 4.0. Right: the centers of 10 groups found by k-means (Lecture 8).

Course map


Part I: basics
Part II: supervised learning
1 Intro
2 Data prep
3 kNN
4 Naive Bayes
5 Trees + forests
6 SVM
7 Boosting
Part III: unsupervised learning
Part IV: deep learning
8 Clustering
9 Neural nets
10 CNNs
11 Trans-formers
12 Diffusion

Part II: supervised learning



weather forecasts

Great game, great fun.

positive

Boring game.

negative

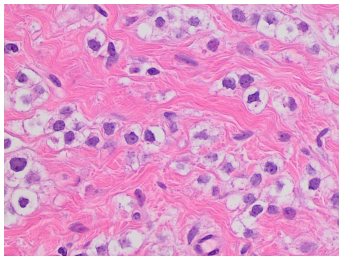
positive or negative reviews



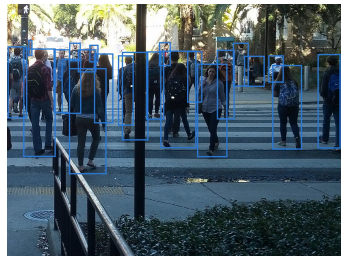
will a customer say yes?

Weather: Thai Meteorological Department (course data). Reviews: Lecture 2. Photo (Wikimedia Commons, cropped): Petiatil, public domain.

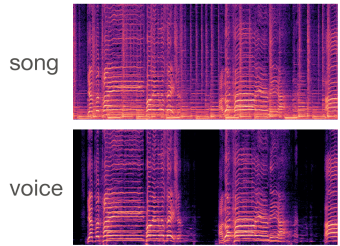
More supervised learning



malignant or benign tumors



self-driving cars



separating the voice from a song

Photos (Wikimedia Commons, cropped): Mikael Häggström, CC0; Erica Fischer, CC BY 2.0, boxes by a pretrained Faster R-CNN (torchvision). Song: “Love in Vain Blues”, Colin Ross Band, CC BY-SA 3.0; voice separated by Hybrid Demucs (torchaudio).

Part III: unsupervised learning

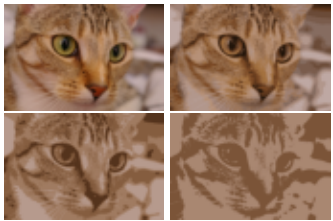
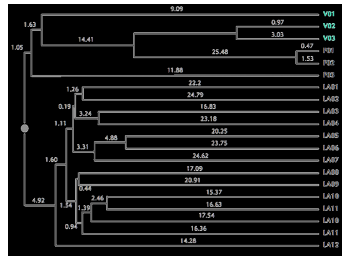


image compression



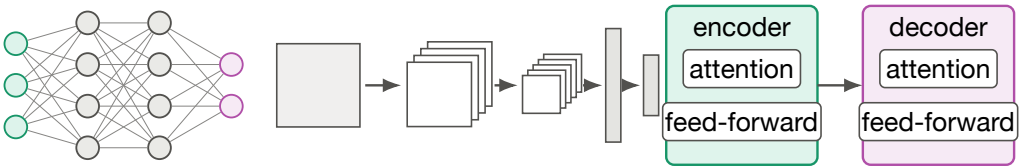
virus family trees



unusual card payments

Cat: scikit-image `chelsea`, CC0 (S. van der Walt), with 16, 4 and 2 colors (Lecture 8). Tree: Metzker et al., PNAS 99 (2002), 14292–14297 (Lecture 8). Card: Mañico, CC0, Wikimedia Commons (cropped).

Part IV: deep learning



feedforward network

Lecture 9

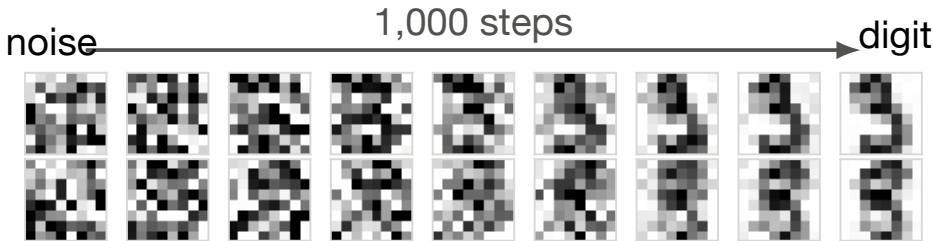
CNN

Lecture 10

transformer

Lecture 11

Images from noise



two samples of our model (Lecture 12)