
# Introduction to Machine Learning

## Course Logistics

林彥宇 教授
Yen-Yu Lin, Professor

國立陽明交通大學 資訊工程學系
Computer Science, National Yang Ming Chiao Tung University

# About Yen-Yu Lin

- Work Experience
  - Professor, *CS, NCTU/NYCU,* August 2019 ~ present
  - Associate research fellow, *CITI, Academia Sinica*, 2015 ~ 2019
  - Assistant research fellow, *CITI, Academia Sinica*, 2011 ~ 2015

- Research interests
  - Computer Vision (**CV**):

    *Let computers see, recognize, and interpret the world like humans*
  - Machine Learning (**ML**):

    *Provide a statistical way to learn how human visual system works*
  - Goal: *Design **ML** methods to facilitate **CV** applications*

# Today's agenda

- Course logistics
- Course overview

# Today's agenda

- Course logistics
- Course overview

# How to choose and take this course?

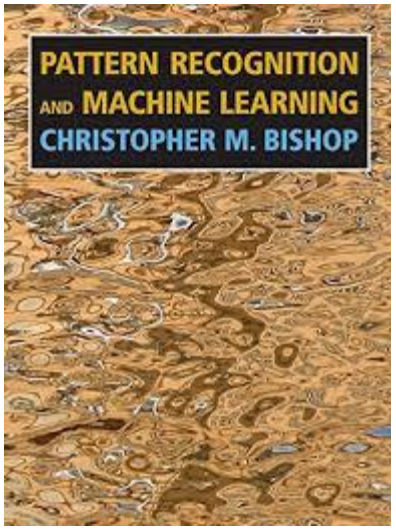
- Please use the online course management system
  - Max number: 90 -> 110 students
- I do not plan to add additional students
  - The size of the classroom
  - The loading of our TAs
  - Teaching quality
  - If you have some strong reason why you must take this course, send me an email with the reason
- Be a guest student?
  - Yes. Send TAs an email with your student ID. We will add you to the student list on E3

# Instructor and teaching assistants

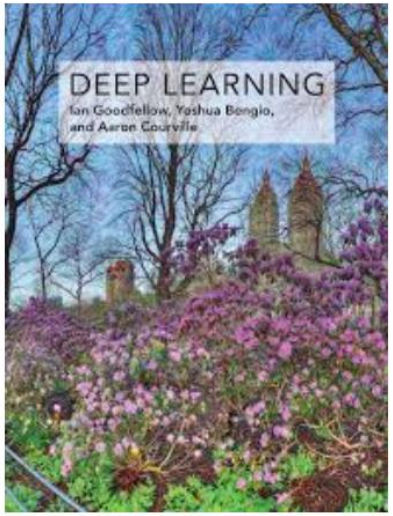
- Instructor: Yen-Yu Lin 林彥宇
  - Email: lin@cs.nycu.edu.tw
  - Office: EC706 (please email me first)
- Teaching assistants:
  - Wei-Hsiang Yu 游為翔 Email: weihsiang.yu@gmail.com
  - Si-Yu Huang 黃思瑜 Email: stella900604@gmail.com
  - Jian-Zhe Wang 王健哲 Email: jzwang.cs13@nycu.edu.tw
  - Chun-Cheng Chu 朱峻正 Email: jasonchu.cs13@nycu.edu.tw
- Office hour (email first)
  - 4:20 pm ~ 5:20 pm on Tuesdays at EC234-C or EC701

# Textbook

- Pattern Recognition and Machine Learning
  - Christopher Bishop
  - Springer-Verlag, Berlin, 2006
  - Free online at
    https://www.microsoft.com/en-us/research/uploads/prod/2006/01/Bishop-Pattern-Recognition-and-Machine-Learning-2006.pdf




- Deep learning (optional)
  - I. Goodfellow, Y. Bengio, and A. Courville
  - MIT Press, 2016
  - Free online at
    https://www.deeplearningbook.org/



# Grading policy

- Four homework assignments: 60% (= 15% x 4)
- For each assignment
  - You are required to implement machine learning algorithms and complete some short answer questions
  - Late policy: 20% off per late day
- Final exam: 40%
  - In the 15th week

# Syllabus

| Week | Date | Course Progress, Contents, Topics |
|---|---|---|
| 1 | 2024-09-03(二) | Introduction |
| 2 | 2024-09-10(二) | Linear Model for Regression I |
| 3 | 2024-09-17(二) | No Lecture: The Mid-Autumn Festival (holiday) |
| HW1 4 | 2024-09-24(二) | Linear Model for Regression II |
| 5 | 2024-10-01(二) | Linear Model for Classification I (pre-recorded videos) |
| 6 | 2024-10-08(二) | Linear Model for Classification II |
| HW2 7 | 2024-10-15(二) | Neural Networks |
| 8 | 2024-10-22(二) | Ensemble Model I |
| 9 | 2024-10-29(二) | Ensemble Model II |
| HW3 10 | 2024-11-05(二) | Kernel Method I |
| 11 | 2024-11-12(二) | Kernel Method II |
| 12 | 2024-11-19(二) | Deep Neural Networks (DNN) |
| HW4 13 | 2024-11-26(二) | Convolutional Neural Networks (CNN) I |
| 14 | 2024-12-03(二) | Convolutional Neural Networks (CNN) II and Transformers I |
| 15 | 2024-12-10(二) | Final Exam |
| 16 | 2024-12-17(二) | Transformers II |

# Pre-requisite

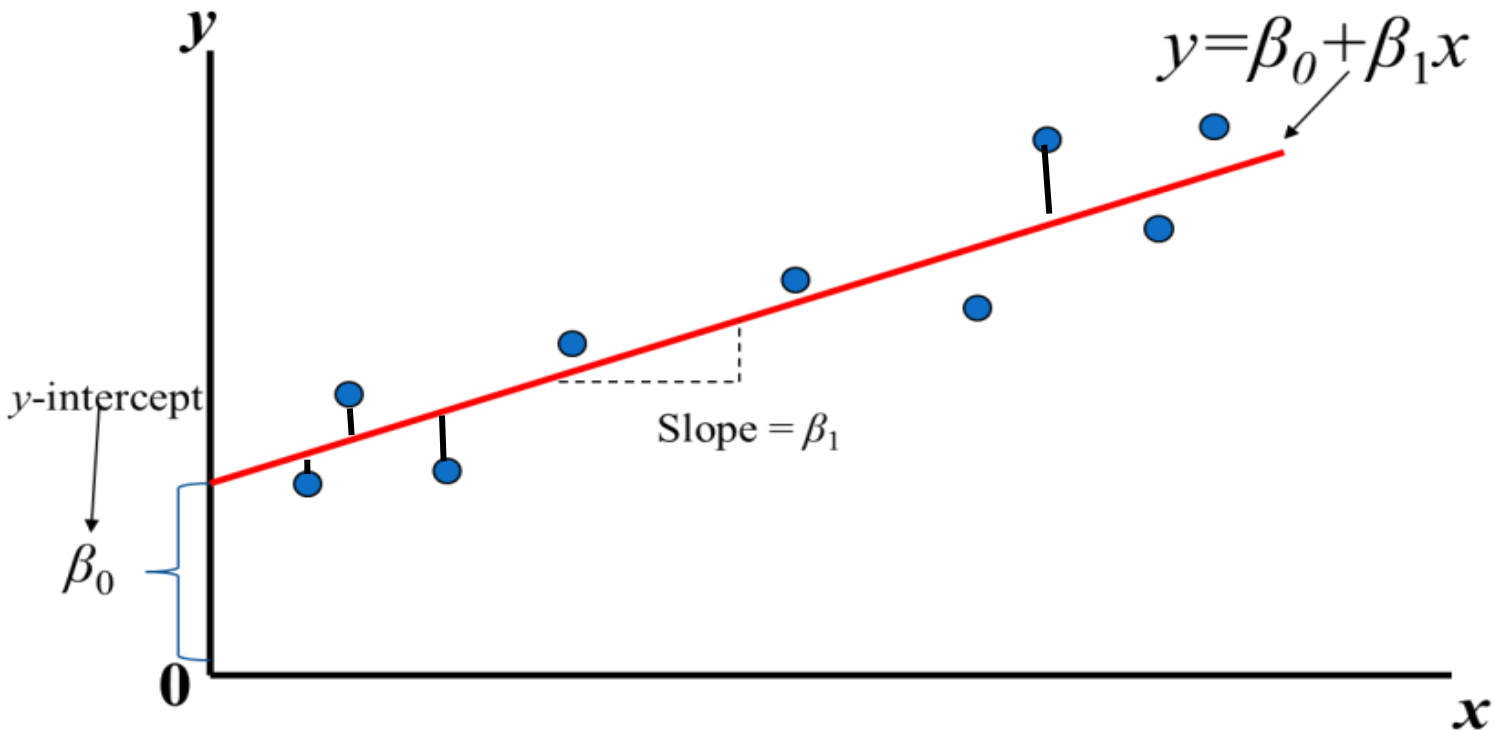
- Linear algebra, probability, calculus, and programming

- Python
  - We strongly encourage students who are not familiar with Python to complete the following tutorial first
  - http://cs231n.github.io/python-numpy-tutorial/

- One deep learning framework, Pytorch or Keras
  - Pytorch: https://pytorch.org/tutorials/
  - Keras: https://elitedatascience.com/keras-tutorial-deep-learning-in-python

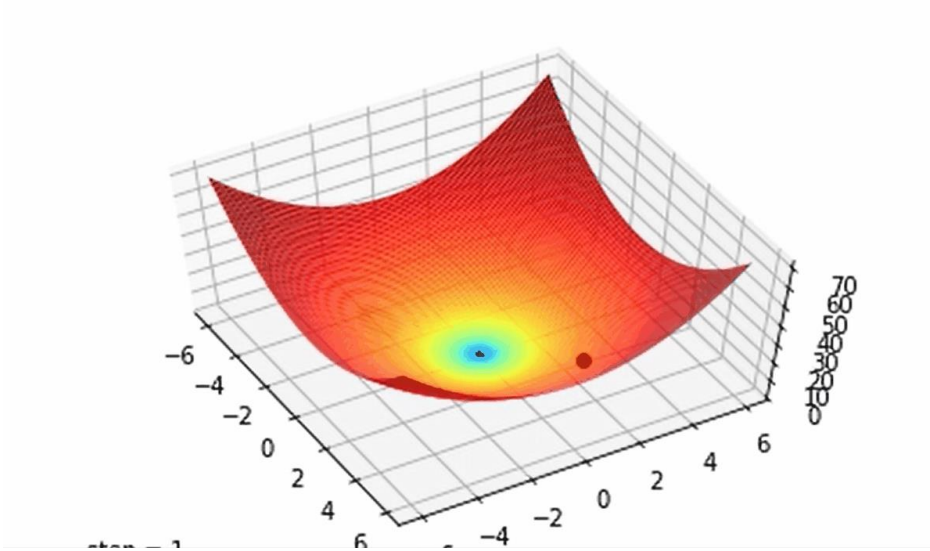
## Homework 1: Linear regression (previous years)

- Find the value of β0 and β1



# Gradient descent

- x-axis and y-axis represent the values of two variables
- z-axis represents the loss of the corresponding variables
- Targets: Find the variable values that minimize the loss



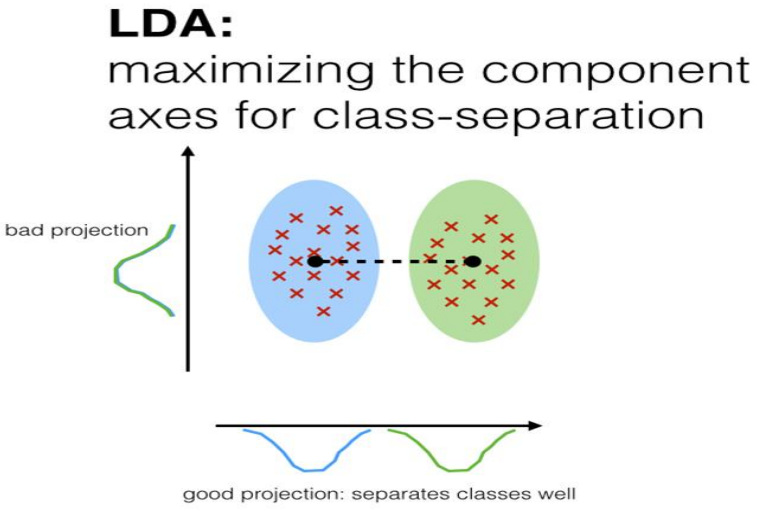
# Gradient descent pseudo code

Algorithm

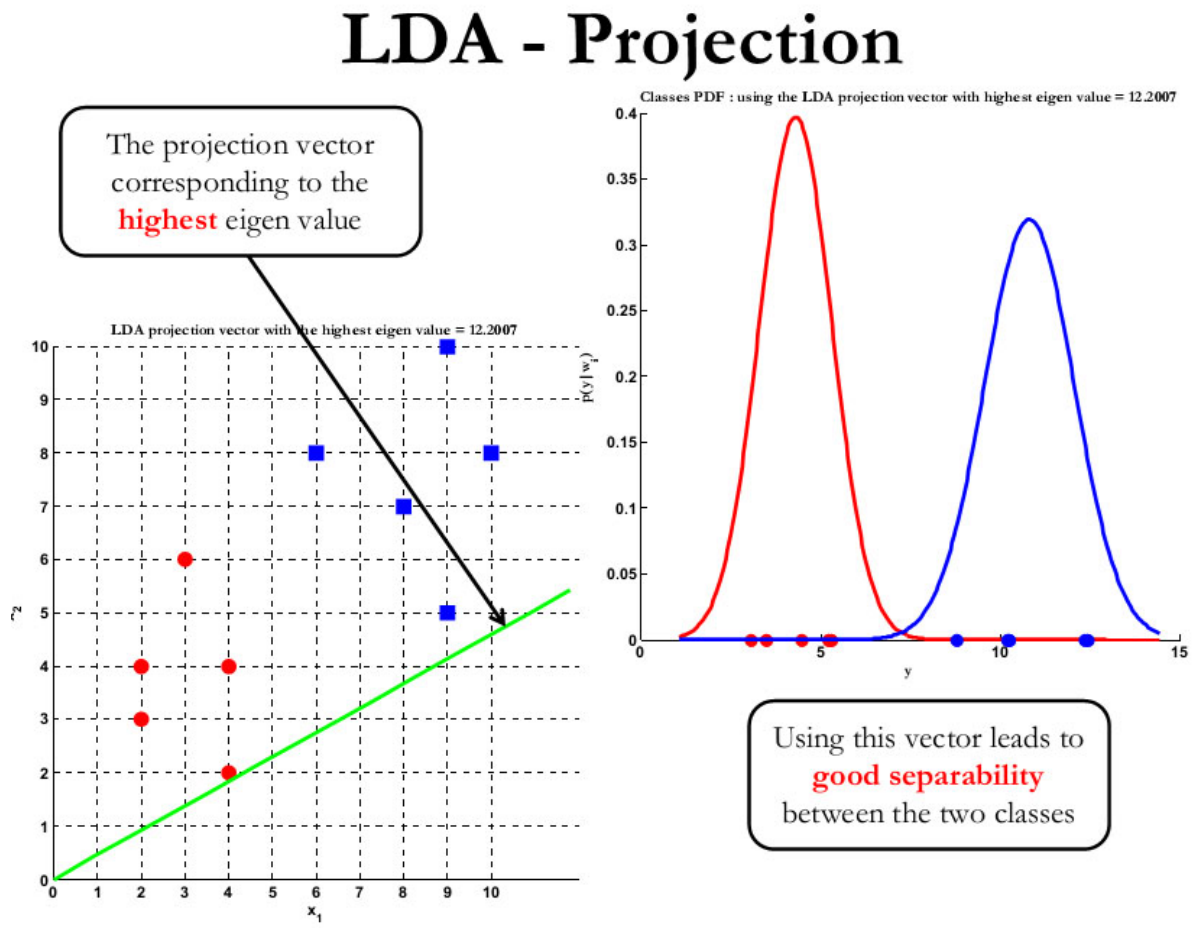
1. Initialize weights randomly $\sim N(0, \sigma^2)$
2. Loop until convergence:
   i. Pick batch of B data points
   ii. Compute gradient. $\frac{\partial J(\Theta)}{\partial \Theta} = \frac{1}{B}\sum_{k=1}^{B} \frac{\partial J_k(\Theta)}{\partial \Theta}$
   iii. Update weights $\theta <- \theta - \eta \frac{\partial J(\Theta)}{\partial \Theta}$
3. Return weights

# Homework 2: Fisher's linear discriminant

- FLD (or LDA) is a "supervised" method and computes the directions representing the axes that maximize the separation between multiple classes.
- FLD seeks the projection **w** that gives a large distance between the projected data means while giving a small variance within each class

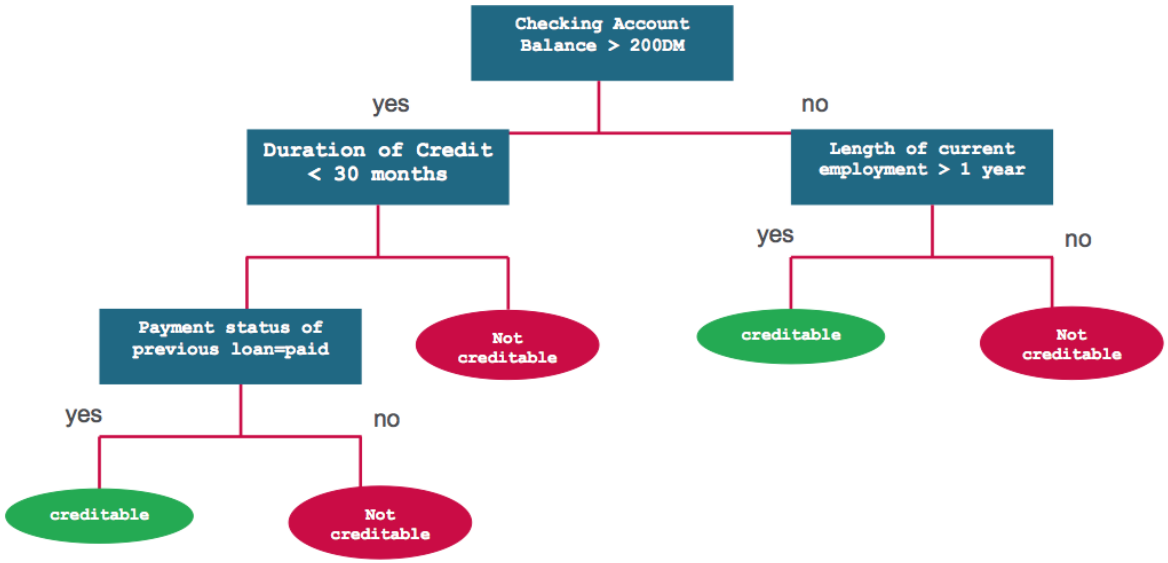


# Eigenvalue problem



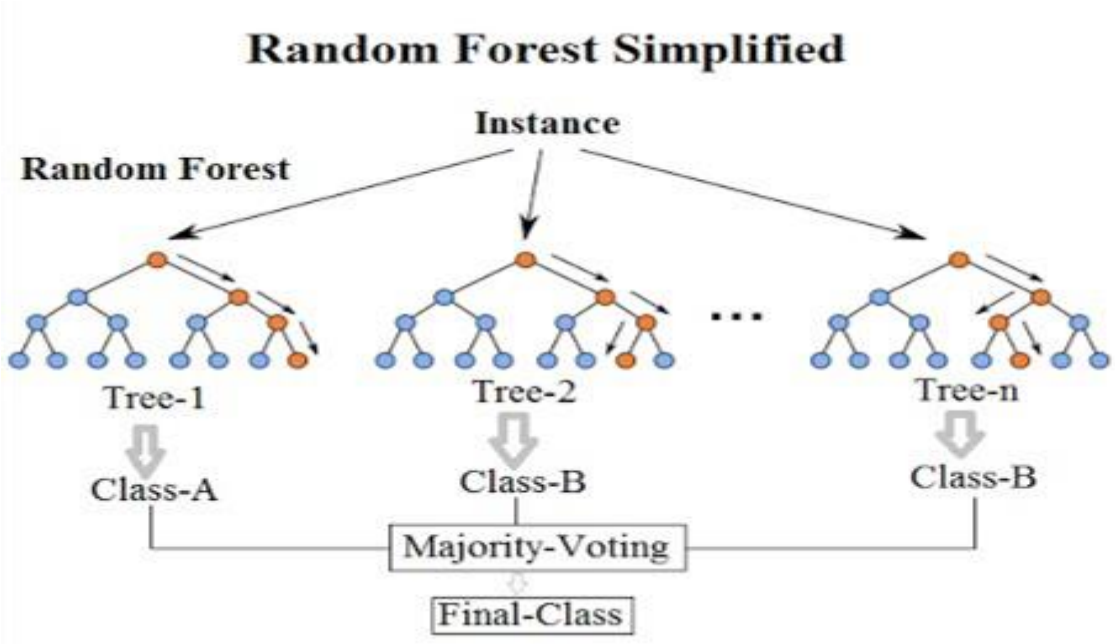
# Homework 3: Decision tree algorithm

- How to find the feature for making decisions? What's the value of feature?
- Find the features to separate data that the class at the resulting nodes are as **pure** as possible



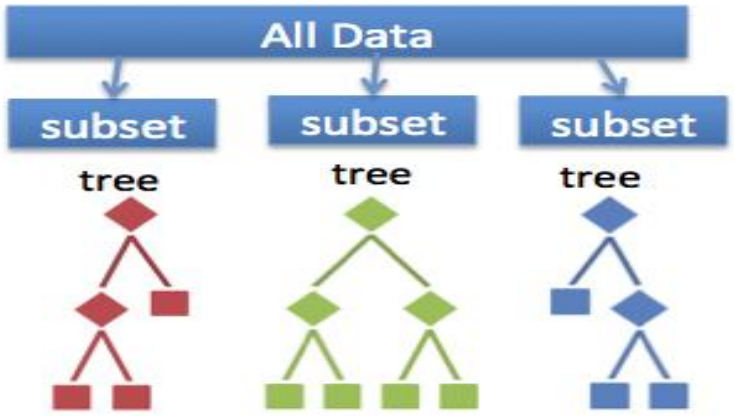
# Ensemble method of decision trees: Bagging

- **Bagging (Bootstrap aggregating)**: Fit many large trees to bootstrap-resampled versions of the training data, and classify by majority vote



# Another ensemble method: Random Forest

- Bootstraped dataset
- Each tree in the forest may grow with different data and features
- Which features or data to be used is **randomly** sampled to grow the tree



# Homework 4: Deep learning

- Use a deep learning algorithm, such as convolutional neural networks or multi-layer perceptron, to carry out image classification or digit detection

# Thank You for Your Attention!

**Yen-Yu Lin (林彥宇)**

Email: lin@cs.nycu.edu.tw

URL: https://www.cs.nycu.edu.tw/members/detail/lin