Machine Learning

Lecture 1: Introduction

Hao Tang

September 22, 2022

https://homepages.inf.ed.ac.uk/htang2/mlg2022/

Face recognition

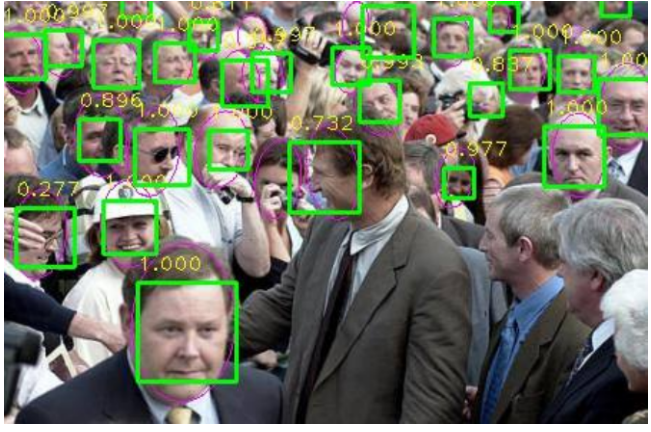


Image credit: (Triantafyllidou and Tefas, 2016)

Photo editing

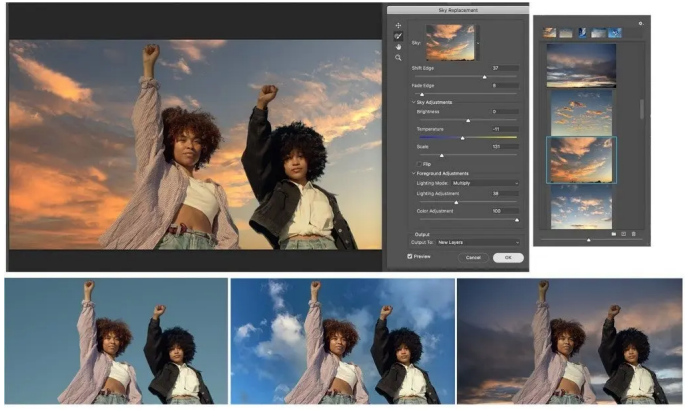


Image credit: https://blog.adobe.com/en/publish/2020/10/20/photoshop-the-worlds-most-advanced-ai-application-for-creatives

Photo editing

Image credit: https://blog.adobe.com/en/publish/2020/10/20/photoshop-the-worlds-most-advanced-ai-application-for-creatives

Video editing

Source | Driver | Ours | Driver | Ours

Image credit: https://samsunglabs.github.io/MegaPortraits/

AlphaGo

Image credit: https://www.theguardian.com/technology/2016/mar/15/alphago-what-does-google-advanced-software-go-next (left)
https://twitter.com/demishassabis/status/692430224382902272 (right)

Recommender systems

NETFLIX

Spotify®

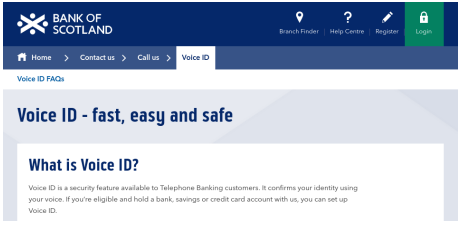
TikTok

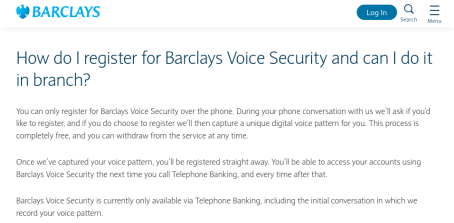
YouTube

Speech recognition

Image credit: https://www.bbc.co.uk/news/technology-47893082

Speech verification





Robotic vacuum cleaner

Image credit: https://www.irobot.co.uk/en_GB/roomba.html

Autonomous driving

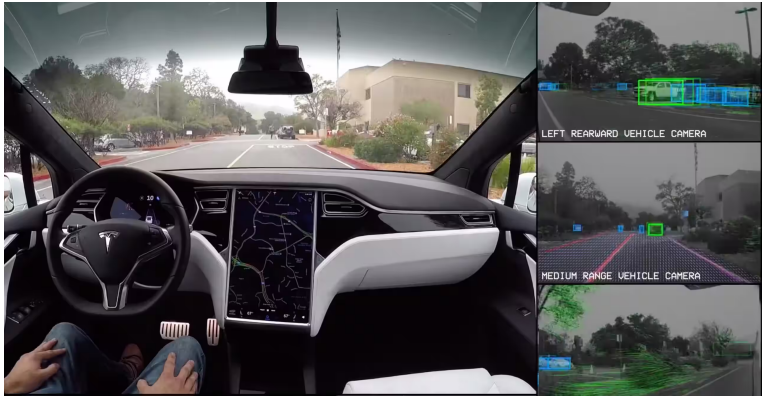


Image credit: https://www.tesla.com/en_GB/videos/autopilot-self-driving-hardware-neighborhood-short

Failures

Image credit: `https://www.nytimes.com/2021/08/17/business/tesla-autopilot-accident.html`

What is machine learning?

What is machine learning?

How would you write a program to recognize hand-written 2s?

What is machine learning?

How would you write a program to recognize hand-written 2s?

What is machine learning?

How would you write a program to recognize hand-written 2s?

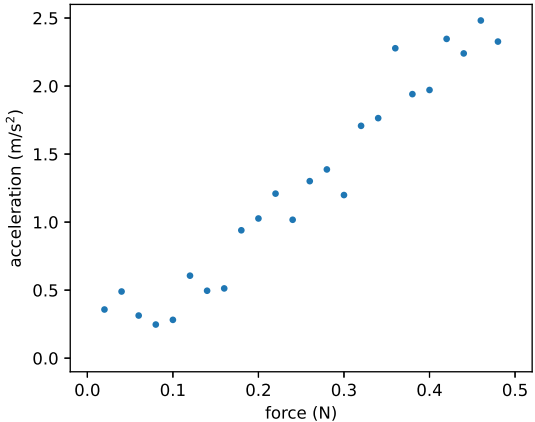
We don't write a program to enumerate all the possible ways of writing 2s.
We "produce" a program using the examples of 2s.

Programming with Data

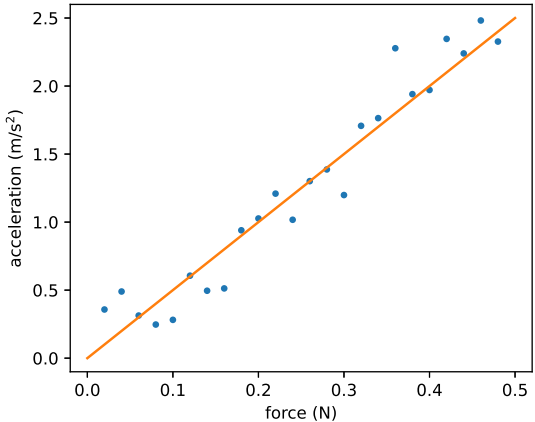
First example

force (N)	acceleration (m/s^2)
0.02	0.358
0.04	0.490
0.06	0.313
0.08	0.247
0.10	0.282
0.12	0.606

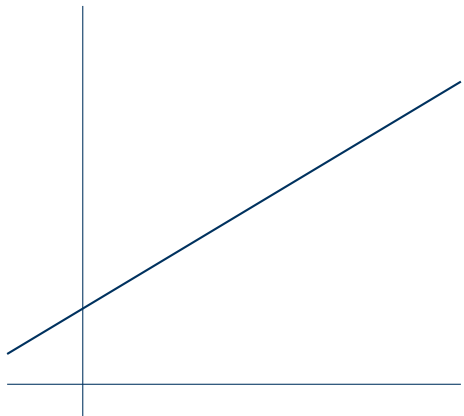
First example



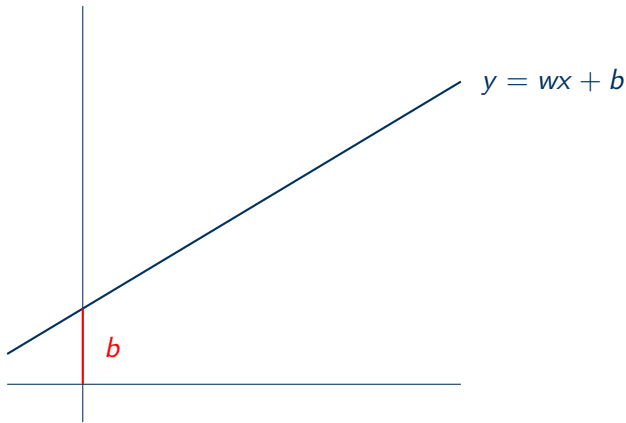
First example



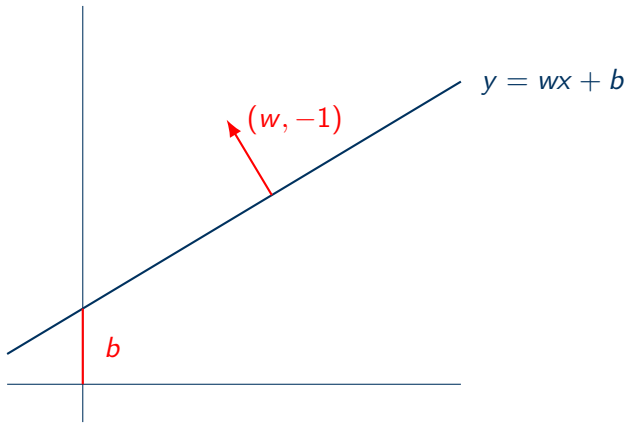
Geometry



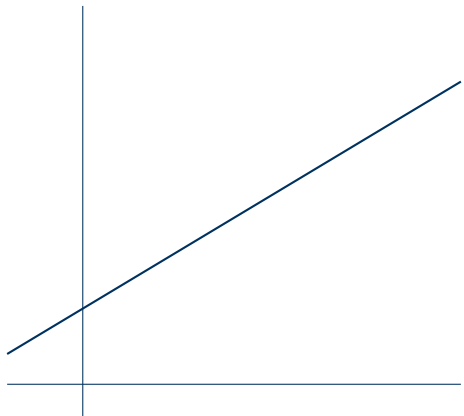
Geometry



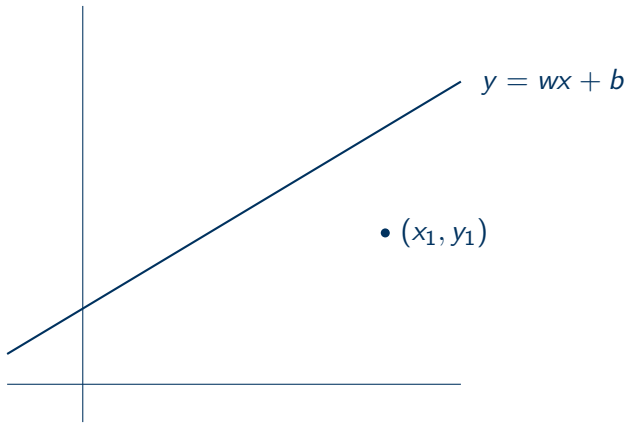
Geometry



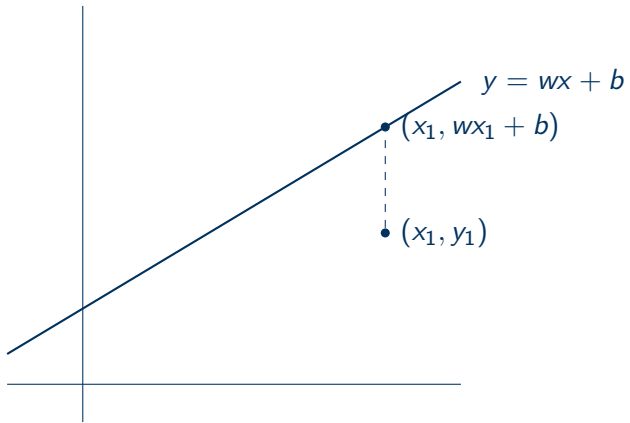
Geometry



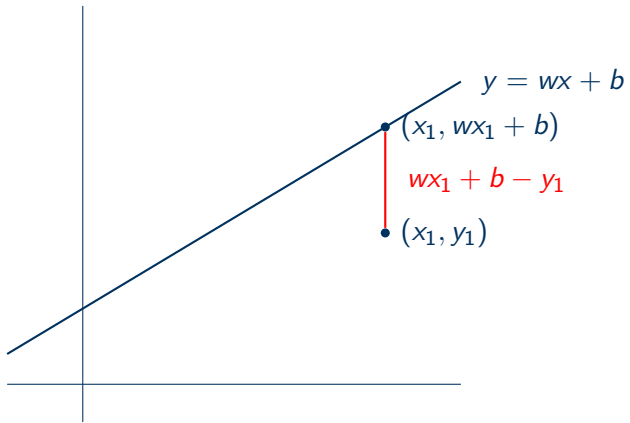
Geometry



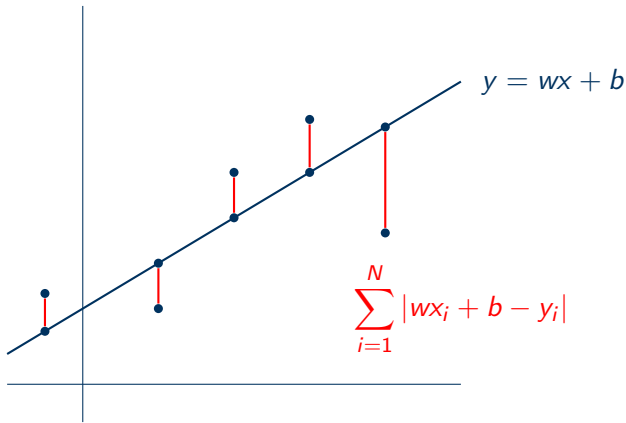
Geometry



Geometry



Geometry



Given N points $\{(x_1, y_1), (x_2, y_2), \ldots, (x_N, y_N)\}$,

$$\text{sum of absolute errors} \qquad L = \sum_{i=1}^{N} |wx_i + b - y_i|. \tag{1}$$

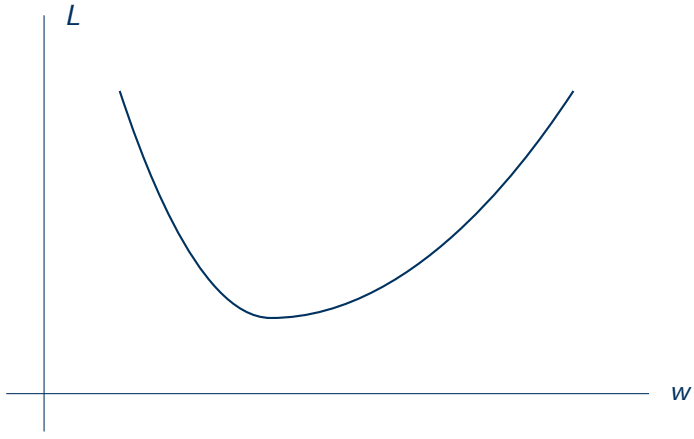
The line that minimizes the error would be

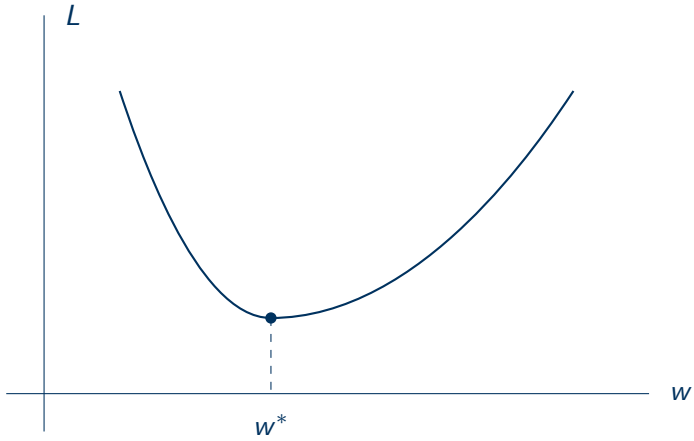
$$\hat{w}, \hat{b} = \underset{w,b}{\operatorname{argmin}}\, L. \tag{2}$$

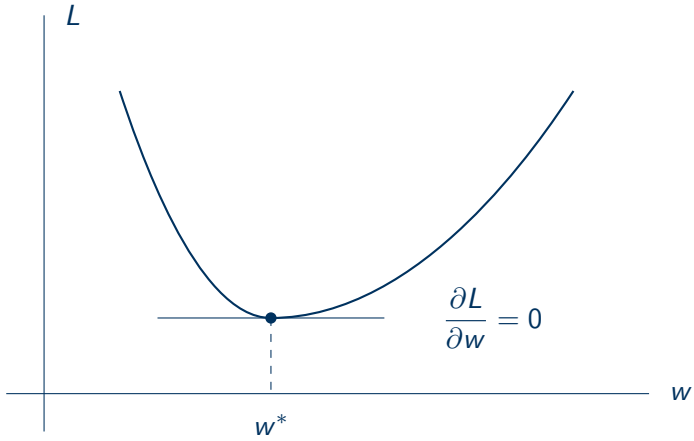
Optimization

How do we solve

$$\min_{w,b} \quad \sum_{i=1}^{N} |wx_i + b - y_i|? \tag{3}$$


L
w


L
w
w*


L
$\frac{\partial L}{\partial w} = 0$
w
w^*

- The problem is that $|a|$ is not differentiable for all $a \in \mathbb{R}$.
- Instead, we modify the loss function to

$$L = \sum_{i=1}^{N} (wx_i + b - y_i)^2. \tag{4}$$

- The plan is to solve

$$\frac{\partial}{\partial w} L = 0 \qquad \frac{\partial}{\partial b} L = 0 \tag{5}$$

$$\frac{\partial}{\partial b}\sum_{i=1}^{N}(wx_i + b - y_i)^2 = \sum_{i=1}^{N} 2(wx_i + b - y_i) \tag{6}$$

$$= 2\left(w\sum_{i=1}^{N} x_i + nb - \sum_{i=1}^{N} y_i\right) = 0 \tag{7}$$

$$b = \frac{1}{n}\sum_{i=1}^{N} y_i - w\frac{1}{n}\sum_{i=1}^{n} x_i = \bar{y} - w\bar{x} \tag{8}$$

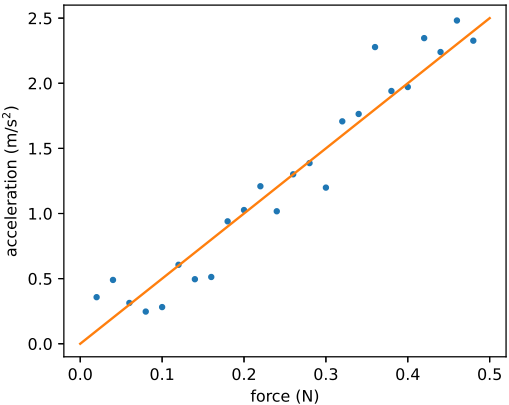
$$L = \sum_{i=1}^{N} (w(x_i - \bar{x}) - (y_i - \bar{y}))^2 \tag{9}$$

$$\frac{\partial}{\partial w} L = \sum_{i=1}^{N} 2(w(x_i - \bar{x}) - (y_i - \bar{y})(x_i - \bar{x}) \tag{10}$$

$$= 2 \left(w \sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x}) - \sum_{i=1}^{N} (y_i - \bar{y})(x_i - \bar{x}) \right) = 0 \tag{11}$$

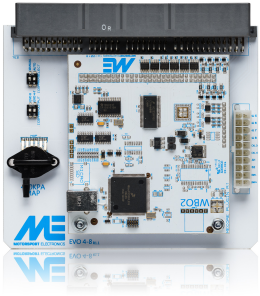
$$w = \frac{\sum_{i=1}^{N} (y_i - \bar{y})(x_i - \bar{x})}{\sum_{i=1}^{N} (x_i - \bar{x})(x_i - \bar{x})} \tag{12}$$

First example




```
def acc(f):
    m = 0.2
    return f / m
```

Temperature calibration



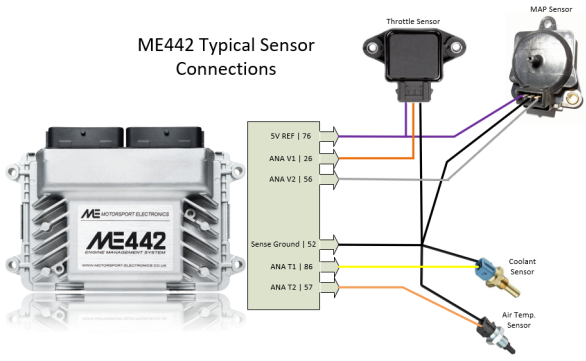
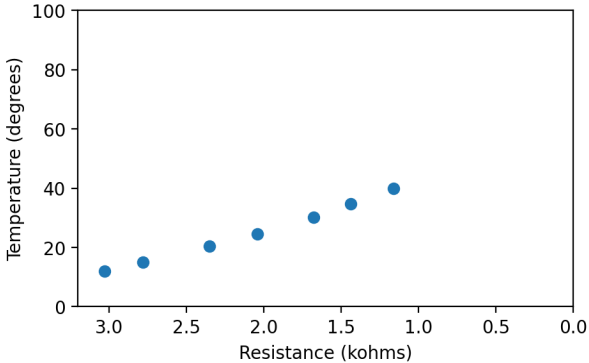
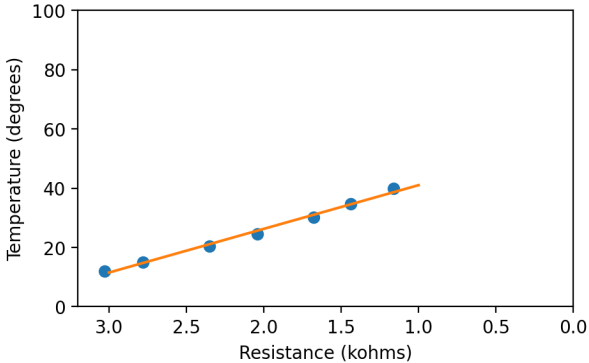


Image credit: https://motorsport-electronics.co.uk/onlinehelp/html/Introduction.html

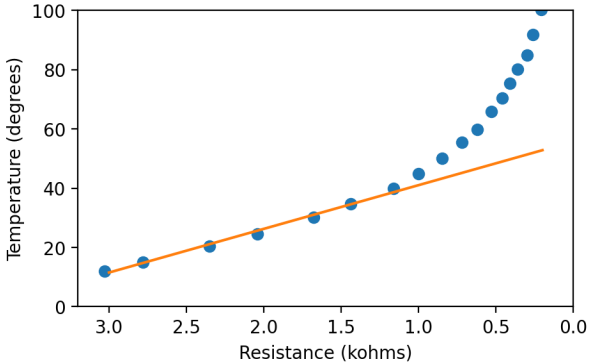
Temperature calibration



Temperature calibration



Temperature calibration



Generalization

- Machine learning models are tools for **prediction**.
- The ability to predict is called **generalization**.
- A learned program is "correct" is it is able to generalize.

Recipe

- Choose a **task**, e.g., classification or regression. (Lecture 5)
- Choose a **data set**. (Lecture 28, 29)
- Choose a **loss**, e.g., mean squared error or cross entropy. (Lecture 4, 5, 6)
- Choose a **model**, e.g., two-layer neural network. (Most of this course)
- Choose an **optimization algorithm**, e.g., stochastic gradient descent. (Lecture 8, 9, 11)
- Check **generalization**. (Lecture 15, 16, 17, 18)

Things we need

- Calculus
- Linear algebra
- Probability
- python, numpy, matplotlib
- ~~Jupiter notebook~~

What you will not learn in this course

What you will not learn in this course

- ~~Use sklearn~~

What you will not learn in this course

- ~~Use sklearn~~
- ~~Stanford CS229: Machine Learning by Andrew Ng~~

What you will not learn in this course

- ~~Use sklearn~~
- ~~Stanford CS229: Machine Learning by Andrew Ng~~
- ~~Kaggle~~

What you will learn in this course

- Formalize a ML problem with math
- Read and understand ML theorems
- Turn ML algorithms into programs

Connections to other courses

- Foundations of Data Science (FDS)
- Applied Machine Learning (AML)
- Machine Learning and Pattern Recognition (MLPR)
- Probabilistic Modeling and Reasoning (PMR)
- Machine Learning Practical (MLP)
- Machine Learning Theory (MLT)
- Reinforcement learning (RL)

Connections to other courses

- Foundations of natural language processing (FNLP)
- Accelerated natural language processing (ANLP)
- Natural language understanding, generation, and machine translation (NLU+)
- Speech processing (in PPLS)
- Automatic speech recognition (ASR)
- Speech synthesis (in PPLS)
- Image and vision computing
- Advanced robotics

Logistics

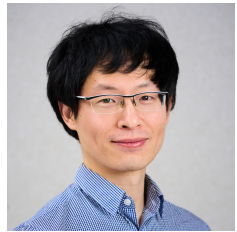
- Course website: `https://homepages.inf.ed.ac.uk/htang2/mlg2022/`
- No textbook. Optional reading on the website.
- 30 hours of lectures
  - Mon at 15:10
  - Thu at 15:10
  - Fri at 15:10
- TA drop-in session Wednesday at 12pm

Logistics

- 5 tutorials on Friday
- 1 practice exam
- 1 coursework (20%)
- 1 final exam (80%)

Lecturers

Kia Nazarpour



Hao Tang

Hiroshi Shimodaira

TAs

Borislav Ikonomov

Sung-Lin Yeh

Ramon Sanabria