



ESCOLA SUPERIOR
NÁUTICA
INFANTE D. HENRIQUE

Machine Learning and Data Science

02. Introduction

Slides based on:
James, G., Witten, D., Hastie, T., & Tibshirani, R. (2023).
An introduction to statistical learning: with Applications in Python.

Outline

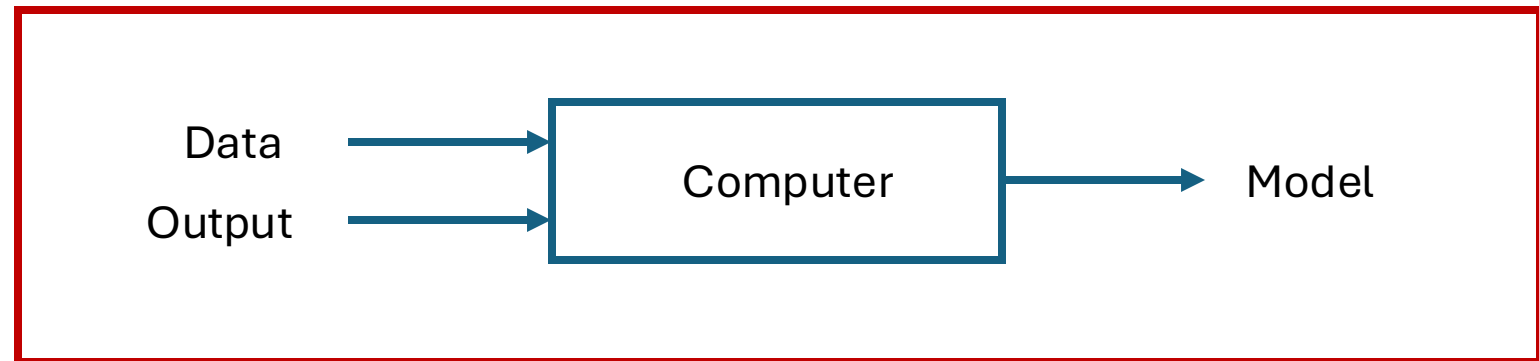
- Motivation
- Some problems
- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

Machine Learning

Classic



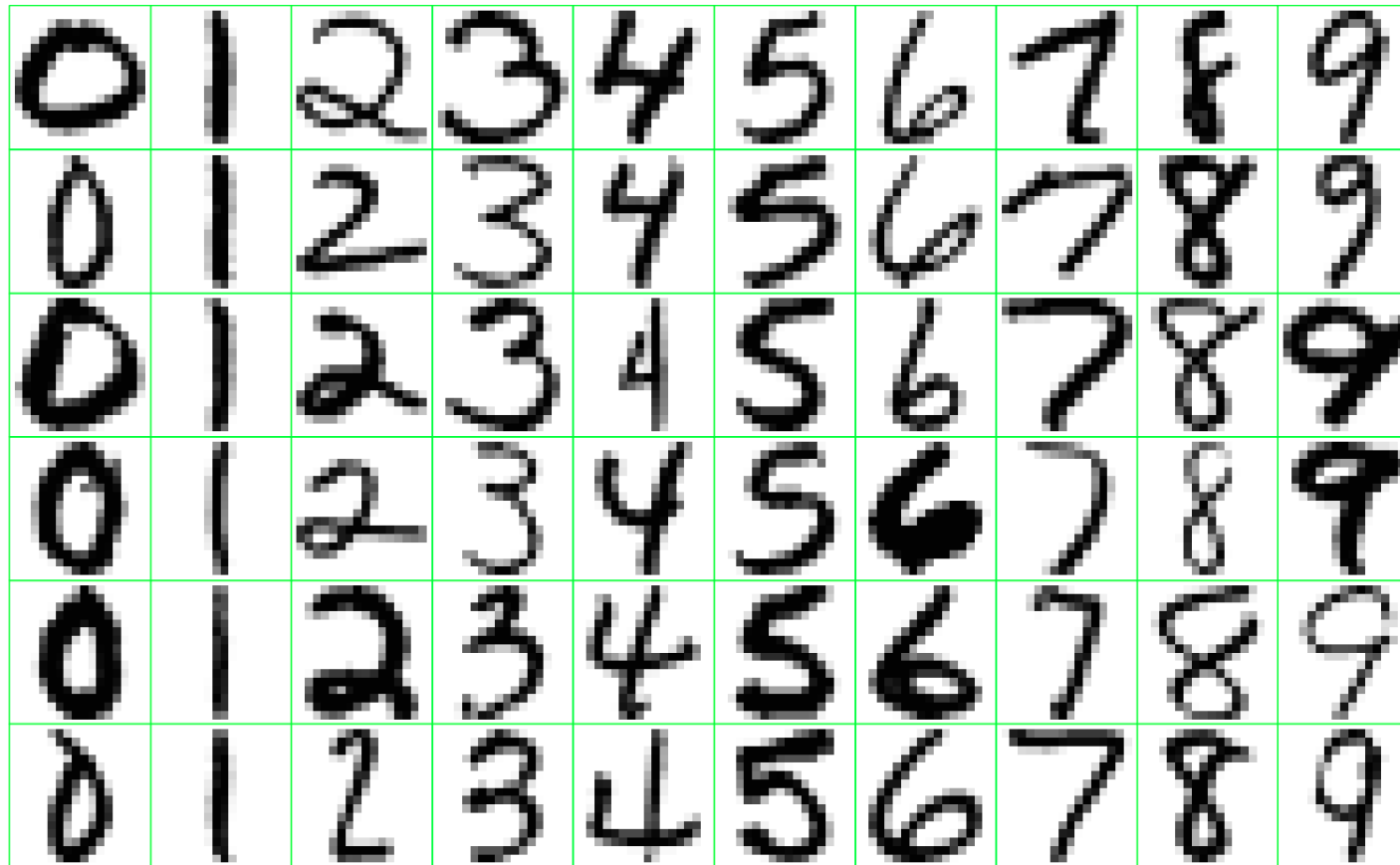
ML



Machine Learning – Some problems

- Risk factors for a given disease
- Predict a disease based on someone's habits
- Spam detection
- Numbers in a handwritten zip code
- Genre of a book
- Movie rating
- ...

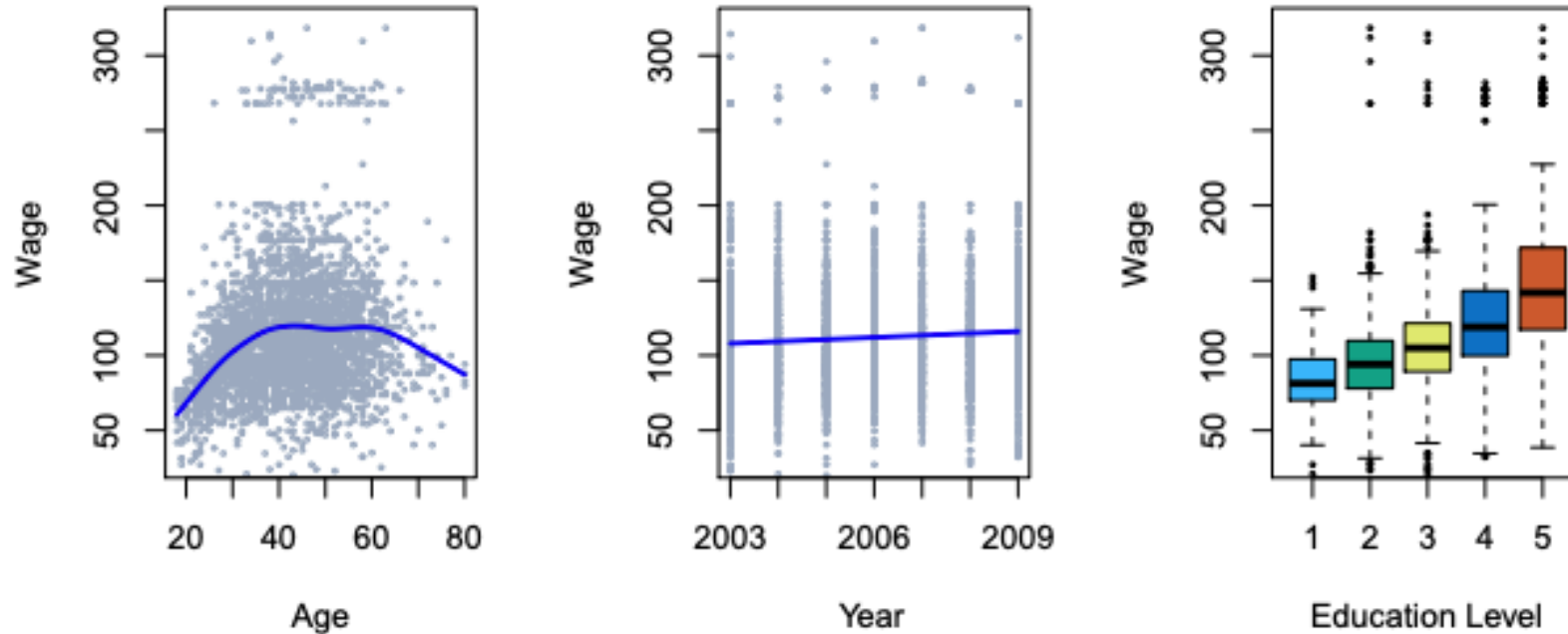
Machine Learning – Some problems



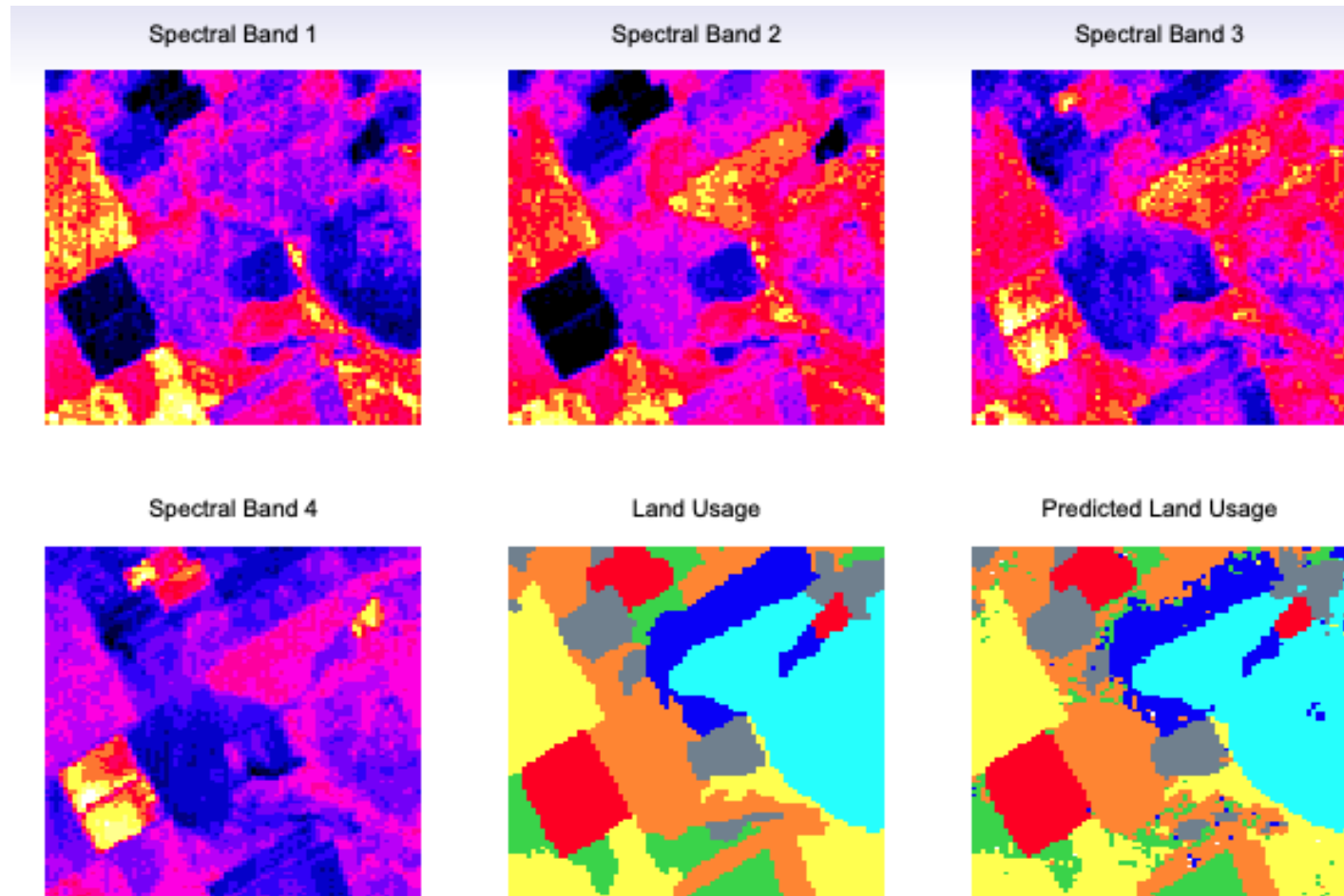
Machine Learning – Some problems

	george	you	hp	free	!	edu	remove
spam	0.00	2.26	0.02	0.52	0.51	0.01	0.28
email	1.27	1.27	0.90	0.07	0.11	0.29	0.01

Machine Learning – Some problems



Machine Learning – Some problems



Machine Learning

- Supervised learning
- Unsupervised learning
- Reinforcement learning

Machine Learning – Supervised Learning

- Output variable: Y (response variable or dependent variable)
- Input variables: $X_1, X_2, X_3, \dots, X_p$ (predictors or features)
- Observations (or instances) : $(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$
- Train/Test
- Predict unseen observations
- Input/output association
- Evaluation

Machine Learning – Unsupervised Learning

- Output variable: ---- (response variable or dependent variable)
- Input variables: $X_1, X_2, X_3, \dots, X_p$ (predictors or features)
- Observations (or instances) : $(x_1), (x_2), \dots, (x_N)$
- Train/Test
- Predict unseen observations
- Input/output association, hidden patterns, similarity, dim reduction, ...
- Evaluation is difficult
- Pre-processing for Supervised Learning

Machine Learning – Reinforcement Learning

- Agent learns to make decisions by interacting with an environment:
 - 1. The agent takes actions (move ahead, turn, etc)
 - 2. Receives feedback (rewards or penalties)
 - 3. Adjusts its strategy to maximize cumulative rewards over time
- Learning: trial and error
- Goal: to develop a policy that tells the agent the best action to take in any given situation to maximize long-term rewards
- There is no labeled dataset
- Learning happens through the agent's experiences

Thank you