

# MA515 - Foundations of Data Science

## Exercise 1

August 6, 2025

- 
1. Which of the following is a classification problem.
    - A. Predicting the salary of a person.
    - B. Predicting the cholesterol level.
    - C. Predicting if a person is heart patient or not.
    - D. Predicting the sales after a advertisement campaign.
  2. **[p.72 ISLR]** For the following parts, indicate whether we would generally expect the performance of a flexible statistical learning method to be better or worse than an inflexible method. Justify your answer.
    - a. The sample size  $n$  is extremely large, and the number of predictors  $p$  is small.
    - b. The number of predictors  $p$  is extremely large, and the number of observations  $n$  is small.
    - c. The relationship between the predictors and response is highly non-linear.
    - d. The variance of the error terms, i.e.  $\sigma^2 = \text{Var}(\epsilon^2)$ , is extremely high.
  3. **[p.73 ISLR]** What are the advantages and disadvantages of a very flexible (versus a less flexible) approach for regression or classification? Under what circumstances might a more flexible approach be preferred to a less flexible approach? When might a less flexible approach be preferred?
  4. From the questions given below, identify if these falls under supervised learning or unsupervised learning. Further, classify them as a regression problem or classification problem.
    - a. Given a list of passengers who survived and did not survive the sinking of the Titanic, predict if someone might survive the disaster.
    - b. Given a set of images of cats and dogs, identify if the next image contains a dog or a cat.
    - c. Given a set of emails on different parameters, labeled as SPAM and Non-SPAM. Identify a new email as spam or not.
    - d. Given a set of movie reviews with sentiment label, identify a new review's sentiment.
    - e. Given images of hand-drawn digit from 0 to 9, identify a number on a hand-drawn digit image.

- f. Given a set of images of the road ahead and degrees to turn the steering wheel. Given a new image of the road ahead, predicting the degree to turn the steering wheel.
- g. Given a set of faces and corresponding names, predict the name of the person in the new image.
- h. Look at a collection of bird photos relying on cues like feather color, size or beak shape, putting them in different groups.
- i. Given GDP growth, CPI, HDI and short term interest rates for 100 countries, identify different groups among them.
- j. Given a set of data on age, education level, country and corresponding salary, predict the salary of a person given age, education level and country.