Chapter 2 - Homework 2, 4, 8, 10

Abdoulfatah Abdillahi

February 2025

1 Exercise 2

Solution: a

- **Type of Problem:** This is a **regression** task because the CEO salary is a continuous numerical variable.
- **Inference vs Prediction:** since we are interested in understanding which **factors** affects CEO salary, which means we want to understand the relationship of the data. So we are primarily interested in **inference**
- **Values of n and p:**
  - $n = 500$ (since we have data on 500 firms).
  - $p = 3$ (since we have three predictors: profit, number of employees, and industry).

Solution: b

- **Type of Problem:** This is a **classification** task because we want to know whether it will be **success** or **failure** for the new product.
- **Inference vs Prediction:** This problem is more toward **prediction** because we want to know whether the new product succeeds.
- **Values of n and p:**
  - $n = 20$ (because we have data on 20 previously launched products).
  - $p = 13$ (because we have 13 predictor variables (price charged, marketing budget, competition price, and ten other variables)

Solution: c

- **Type of Problem:** This is a **regression** task because we are predicting a continuous variable: the
- **Inference vs. Prediction:** This problem is more focused on **prediction** because we are interested in forecasting the

- **Values of n and p:**
  - n = *Weekly data for all 2012.*
  - $p = 3$ (the % change in the US market, the British market, and the German market).

2 Exercise 4

Three Real-life Applications in Classification:

- **1. Alzheimer's Disease Classification:** I worked on an Alzheimer's Disease classification project using PET images. The goal was to classify patients into three categories: Cognitively Normal (CN), Early Mild Cognitive Impairment (EMCI), and Alzheimer's Disease (AD). This is a real-life application of classification because it helps in predicting the patient's condition, allowing for early diagnosis and better treatment planning. Since the primary objective is to assign patients to predefined categories, this is a classification task focused on prediction.
- **2. Antibiotic Resistance Prediction in E. coli:** My current project focuses on predicting antibiotic resistance in *E. coli* using deep learning models. The goal is to classify whether a given drug is resistant or susceptible. This is a real-life application of classification because we are developing a model that helps determine whether a patient's bacterial infection is resistant to a specific antibiotic. Since the objective is to assign a binary label (**resistant or susceptible**), this is a classification task primarily focused on prediction.
- **3. Fraud Detection:** Classifying whether a financial transaction is fraudulent (Fraud/No Fraud) based on the transaction amount, location, and user behavior. This is a real-world application that detects fraudulent activity, and the goal is to understand which transaction patterns or behaviors are most indicative of fraud. Therefore, this is a classification task focused on inference.

Three Real-life Applications for Regression:

- **1. Predicting House Prices:** Real estate companies use multiple linear regression to predict house prices based on various factors such as square footage, number of bedrooms, and location. By analyzing these variables, they can estimate the market value of a house. The aim of this task is to predict the value of the house.
- **2. Medical Research:** Medical research often uses linear regression to understand the relationship between drug dosage and patient outcomes.

For example, in my research, I am using linear regression to understand the relationship between patient characteristics and drug use. By analyzing the data, I can infer the optimal dosage that minimizes side effects while maximizing therapeutic benefits. This task's aim is inference because researchers use regression to understand the relationship between variables.

- **3. Marketing Analysis:** Many businesses use regression analysis to predict the impact of advertising spending on sales revenues. By examining historical data, they can determine the effectiveness of different advertising channels and allocate their budget more efficiently.

Three Real-life Applications for Cluster:

- **1. Customer Segmentation in Marketing:** Companies use clustering to segment their customers into distinct groups based on purchasing behavior, demographics, and other attributes. For instance, Amazon recommends products similar to those a customer is purchasing. This helps businesses tailor marketing strategies and personalized campaigns for different customer segments, ultimately improving customer satisfaction and sales.

  This is a clustering task because we are grouping customers based on their behavior without predefined categories. The aim is inference, as the goal is to understand customer behavior and identify distinct customer groups rather than making direct predictions.

- **2. Anomaly Detection in Network Security:** Clustering can be used to identify unusual patterns in network traffic that may indicate security threats or breaches. By grouping normal behavior patterns, any deviation from these clusters can be flagged as a potential anomaly, helping with the early detection and prevention of cyber-attacks.

  This is a clustering task because we are identifying patterns without predefined labels and detecting outliers based on deviations from established groups. The aim is inference, as the goal is to understand deviations from normal behavior rather than predicting specific future events.

- **3. Image Segmentation in Medical Imaging:** In medical imaging, clustering algorithms are used to segment different regions of an image, such as identifying tumors or other abnormalities in MRI or CT scans. This helps radiologists and medical professionals to analyze and diagnose medical conditions more accurately and efficiently. This application if more focused on inference.

3 Exercise 8 and 10 : Check the google colab, I attached another file