Exercises

1 Probabilities

Given a Doctor's Patient Record, where each cell $n(x_r, \theta_c)$ represents the number of patients with symptom x_r and disease θ_c:

	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6	θ_7	θ_8	θ_9	θ_{10}	**Sum**
x_1	8	9	9	5	4	1	1	0	0	0	37
x_2	3	5	8	9	14	10	3	3	0	0	55
x_3	0	1	1	10	16	11	12	7	8	5	71
x_4	0	0	1	0	3	5	10	7	7	4	37
Sum	11	15	19	24	37	27	26	17	15	9	200

Table 1: Joint observations of symptoms x_r (rows) and diseases θ_c (columns).

Questions from the doctor:

- **Q1:** What is the joint probability $P(x_3, \theta_2)$ that a patient has the symptom x_3 and the disease θ_2?
- **Q2:** What is the probability $P(x_3)$ that a patient has the symptom x_3?
- **Q3:** What is the probability $P(\theta_2)$ that a patient has the disease θ_2?
- **Q4:** What is the conditional probability $P(x_3|\theta_2)$ that a patient has the symptom x_3 given that they have the disease θ_2?
- **Q5:** What is the conditional probability $P(\theta_2|x_3)$ that a patient has the disease θ_2 given that they have the symptom x_3?

2 Maximum Likelihood

1. We flipped a coin 100 times. Given that there were 55 heads, use maximum likelihood estimation to find the probability p of heads on a single toss.
2. Let X be independent and identically distributed (i.i.d.) Poisson(λ) distributed. Find the maximum likelihood estimator for λ, $\hat{\lambda}$. Calculate an estimate using this estimator when:

$$x_1 = 1, x_2 = 2, x_3 = 5, x_4 = 3.$$

3 Logistic Regression

A company wants to predict whether a customer will purchase a product based on their income and age. The logistic regression model is given by:

$$P(Y = 1|X) = \frac{1}{1 + e^{-(\beta_0+\beta_1 X_1+\beta_2 X_2)}}$$

where:

- $Y = 1$ if the customer buys the product, $Y = 0$ otherwise.
- X_1 = Age of the customer.
- X_2 = Monthly income of the customer (in £1000s).
- $\beta_0 = -3$, $\beta_1 = 0.05$, $\beta_2 = 0.1$.

Tasks:

1. Compute the probability that a customer aged 30 with a monthly income of £5000 will purchase the product.
2. Interpret the meaning of $\beta_1 = 0.05$ in the context of this problem.
3. If a customer's probability of purchasing is 0.7, should the company predict that they will buy the product? Justify your answer.
4. Briefly explain one advantage and one limitation of using logistic regression for this problem.

4 Naïve Bayes - Disease Diagnosis

A medical test is used to diagnose a rare disease based on two symptoms: Fever and Cough. The model uses a Naïve Bayes classifier.
The probabilities of symptoms appearing given the disease status are:

Symptom	P(symptom|disease)	P(symptom|no disease)
Fever	0.8	0.2
Cough	0.6	0.3

Additionally, we know that only 5% of the population has the disease (i.e., $P(\text{disease}) = 0.05$).
Tasks:

1. Compute the probability that a patient with both Fever and Cough has the disease, and classify the patient as having or not having the disease.
2. What is the 'naive Bayes' assumption and why is it important in Naïve Bayes classification?