Introduction to Machine Learning, Spring 2022
Homework 2

(Due Friday, Apr. 8 at 11:59pm (CST))

March 24, 2022

1 Problem1

Given a set of data $\{(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)\}$, where $y_i \in \{0, 1\}$. We want to conduct a binary classification, and the decision boundary is $\beta_0 + x^T\beta = 0$. When $\beta_0 + x^T\beta > 0$, the sample will be classified as 1, and 0 otherwise.

(a) Define a function which enables to map the range of an arbitrary linear function to the range of a probability [2 points]

(b) Derive the posterior probability of $P(y_i = 1|x_i)$ and $P(y_i = 0|x_i)$ [3 points]

(c) Write the log-likelihood for N observations, which means:

$$l(\theta) = logP(Y|X) = \sum_{i=1}^{N} log(P(y_i|x_i))$$

(Using the expression of $P(y_i|x_i)$ in (b) and eliminate redundant items) [5 points]

Table 1: probability distribution for X

X	0	1	2	3
P	θ^2	$2\theta(1-\theta)$	θ^2	$1-2\theta$

2 Problem2

(a) Given a random variable X and its probability distribution is shown in Table 1. Now, we sample 8 times and get the results $\{3, 1, 3, 0, 3, 1, 2, 3\}$. Please derive the MLE estimate for $\theta (0 < \theta < \frac{1}{2})$. [4 points]

(b) Now we discuss Bayesian inference in coin flipping. Let's denote the number of heads and the total number of trials by N_1 and N, respectively. Please derive the MAP estimate based on the following prior:

$$p(\theta) = \begin{cases} 0.5 & \text{if } \theta = 0.5 \\ 0.5 & \text{if } \theta = 0.3 \\ 0 & \text{otherwise,} \end{cases}$$

which believes the coin is fair, or is slightly biased towards tails. [4 points]

(c) Suppose the true parameter is $\theta = 0.31$. Please compare the prior in (b) with the Beta prior distribution(You can review this part in Lecture 07). Which prior leads to a better estimate when N is small? Which prior leads to a better estimate when N is large? [2 points]

3 Problem3

According to the following Fig. 3, answer the following questions:

(a) use the D-separation to discus whether the following statements are true or not:

   (1) Given x_4, $\{x_1, x_2\}$ and $\{x_6, x_7\}$ are conditionally independent.[1(reason)+1(conclusion) points]

   (2) Given x_6, x_3 and x_2 are conditionally independent.[1(reason)+1(conclusion) points]

(b) if all the nodes are observed and boolean variables, please complete the process of learning the parameter $\theta_{x_6|i,j}$ by using **MLE**, where $\theta_{x_6|i,j} = p(x_6 = 1 \mid x_3 = i, x_4 = j), i, j \in \{0, 1\}$.[6 points]

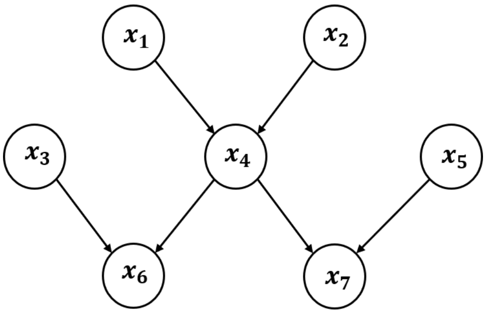

Figure 1: The Bayesian network for questions 3.