



考試日期： 年 月 日

考試科目	資料科學	系所別	人工智慧博士學位學程	命題教師	
<p>1. Please use an example to explain Bayesian Theorem and Bayesian Reasoning. (10%)</p> <p>2. What is Bayesian Classifier? Please give an example and explain how to use Bayesian Classifier to predict spam emails. (10%)</p> <p>3. For data science, the correctness of the data is very important. Therefore, it is necessary to confirm the distribution and integrity of the data through data preprocessing. Please describe the possible works of data preprocessing for data. (20%)</p> <p>4. For the analysis work of data science, the larger the number of features (or independent variables) of numerical data, the results may be bad. The reason may cause overfitting or collinearity. Please explain possible ways to overcome the problem of overfitting or collinearity. (20%)</p> <p>5. Outliers are often discarded as noise. However, some exceptions in credit card transactions might be detected as the fraudulent use of credit cards. Using fraudulence detection as an example, propose three potential methods that can be used to detect outliers. In your opinion, which one is more reliable than the others? Please explain your comparison. (10%)</p> <p>6. Suppose that the instance space $\mathcal{X} = \{0,1\}^d$ and consider the target function c^* that labels an instance $\mathbf{x} = (x_1, x_2, \dots, x_d) \in \mathcal{X}$ as “positive” if the least index i for which $x_i = 1$ is odd, and otherwise, labels \mathbf{x} as “negative”. Show that the decision rule of c^* can be represented by a linear threshold function. (10%)</p> <p>7. Let $a_1, a_2, \dots, a_n, \dots$ be a stream of symbols, each of which is an integer in $\{1, 2, \dots, M\}$ and arrives one at a time. The whole data stream is almost infinite so that it is too large to be stored in random access memory.</p> <p>(a) Give an algorithm that will select a symbol uniformly at random from the stream. Please show the validity of the proposed algorithm. How much memory does your algorithm require? (10%)</p> <p>(b) Give an algorithm that will select a symbol with probability proportional to a_i^2. Please show the validity of the proposed algorithm. (10%)</p>					