人工智能 （B 类）CS41

期末考试

姓名：

学号：

A total of 11 questions. Please chose any of the 10 questions to answer.

Question 1 (10 points)： What is the statistical justification of using Least Square as the objective function during a regression analysis, assuming the training data are all derived from i.i.d. Gaussian models (identically independent distributed)? Please give simple mathematical derivations.

Question 2 (10 points): If the training data of different labels are not perfectly derived from Gaussian distributions (or not sure about what distributions might be best used to model the data), which model would you chose to both distinguish different data while minimize errors? Please give a simple mathematical representation to generalize the idea.

Question 3 (10 points): In Naïve Bayes (NB), what additional

modifications are needed in order to classify more than 2 different types of training data (Multivariate NB)? Please give a simple mathematical form to highlight the difference to the regular NB?

Question 4 (10 points): Why multinomial NB can be used to better describe some contextual information in test mining? Please describe the key mathematical differences between multivariate NB and multinomial NB?

Question 5 (10 points): SVM is a constrained optimization problem. 1) Please use the generalized Lagrangian to represent the problem; 2) Please write down the Primal and Dual problems.

Question 6 (10 points): Given SVM is proved to possess the nature of strong duality, what key additional insights are we able to get? Why only the support vector matters during the learning? Please give the simple mathematical justification.

Question 7 (10 points): Please define the Kernel function with simple mathematical terms. Please give your interpretations of

a distance measure in the context of kernel, and how you might use the concept to distinguished complex models linearly?

Question 8 (10 points): What would be the key principle in machine learning to minimize overfitting and underfitting? Besides data, what would be key issue when selecting model? Please give one technique as an example to select a better model.

Question 9 (10 points): Describe mathematically how PCA is derived as a constrained optimization problem. Please start from data variances to the final solution of principle components.

Question 10 (10 points): What is the key statistical justification of EM algorithm? Please compare to the regular likelihood estimation. However, there might still be problems such as local solutions etc, what would be your solutions when facing these problems？

Question 11 (10 points) For HMM, please discuss the different computational complexities between a naïve approach vs. the

use of Markov condition of independences as well as dynamical programming. Please provide simple mathematical representations.