

THE HONG KONG UNIVERSITY OF SCIENCE & TECHNOLOGY
Machine Learning
Homework 2

Due Date: 18/10/19 (CSIT), 19/10/19 (MSBD)

To submit your work, hand it to the instructor on the due date.

Question 1 Consider the following dataset:

Instance	y	x_1	x_2
1	1	0	0
2	1	0	0
3	1	0	1
4	1	0	1
5	0	1	0
6	0	1	0
7	1	1	1
8	0	1	1

- (a) Give the Naïve Bayes model for the data. There is no need to use Laplace smoothing, and there is no need to show the process of calculation.
- (b) Calculate the posterior probabilities of the Instances 1 and 7 belonging to the two classes according to the model of the previous sub-question. Show the process of calculation.

Question 2 Explain how Gaussian discriminant analysis (GDA) is related to logistic regression. Discuss their pros and cons in terms of the distribution characteristics of data and the amount of data.

Question 3 Discuss other major pros and cons of generative classifiers and discriminative classifiers that are not covered in Question 2.

Question 4 The following figure shows the general trend in the training and test errors of classifiers as a function of sample size.

- Which curve represents the training error? Briefly explain.
- What does the gap between the two curves represent? According to the VC Theorem, how does the gap depend on the sample size and model complexity respectively?

