COMP90049 Introduction to Machine Learning
Semester 1 2021

Assignment 2 feedback for **Kaixun Yang (1040203)**

Criterion	Available marks	Your mark
Question 1a		
Mark	0.5	0.5
Question 1b		
Mark	0.5	0.25
Question 2		
Mark	0.5	0.5
Question 3a		
Mark	1	1
Question 3b		
Mark	1	1
Question 4a		
Mark	1.5	1.5
Question 4b		
Mark	0.5	0.5
Question 4c		
Mark	1	0.5
Question 5a		
Mark	0.5	0.5
Question 5b		
Mark	2	1.5
Question 6		
Mark	1	1
Question 7a		
Mark	1	0.25
Question 7b		
Mark	1	0.25
Question 7c		
Mark	1	1
Question 7d		
Mark	1	1
Question 7e		
Mark	1	0.25

Criterion	Available marks	Your mark
Late Penalty		
Mark		**0**
Overall mark for Assignment 2	**15**	**11.5**

Feedback

Q1b: you should fit your OneHotEncoder on the training set and use the same instance to transform test set (shouldn't use fit_transform for test data, should use `transform`)
Q4a: nb_full implementation is incorrect (see comment/deduction in 4c)
Q4c: There's a simpler solution: multiply the predictions of the Gaussian/Bernoulli NB and multiply them with the prior. This approach is justified by the conditional feature independence assumption of NB
Q5b: you should point out that numeric features get larger coefficients than categorical ones
Q7a: you should explain your observations in terms of mechanisms of the classifier or data used
Q7b: Class imbalance is the central reason; reference to data missing
Q7e: neither is guaranteed perfect performance (e.g., because we can't guarantee that the data is linearly separable);