ANSWER BOOKLET

1. Name: Rohak Singhal

2. Student ID: 3035242475

3. Marks:

Total marks: 100

Weighing: 10% of overall assessment

Question 1: Reading Text files using Octave/Python

Q1. Submit Q1.m/Q1.py (Octave/Python)

(10 marks)

Question 2: Feature Extraction

Q2(a) Submit Q2a.m/Q2a.py (Octave/Python)

(10 marks)

Q2(b) Submit Q2b.m/Q2b.py (Octave/Python)

(10 marks)

Q2(c). Write down the number of occurrence of the string "feature1" (1st keyword) in "01.txt". (5 marks)

feature1: Car

Occurances: 1

Q2(d). Write down the number of occurrence of the string of all keywords in "01.txt". (5 marks)

keywords: ['car', 'passenger', 'seat', 'drive', 'power', 'highway', 'purchase', 'hotel', 'room', 'night', 'staff', 'water', 'location']

Occurances: [1, 1, 1, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0]

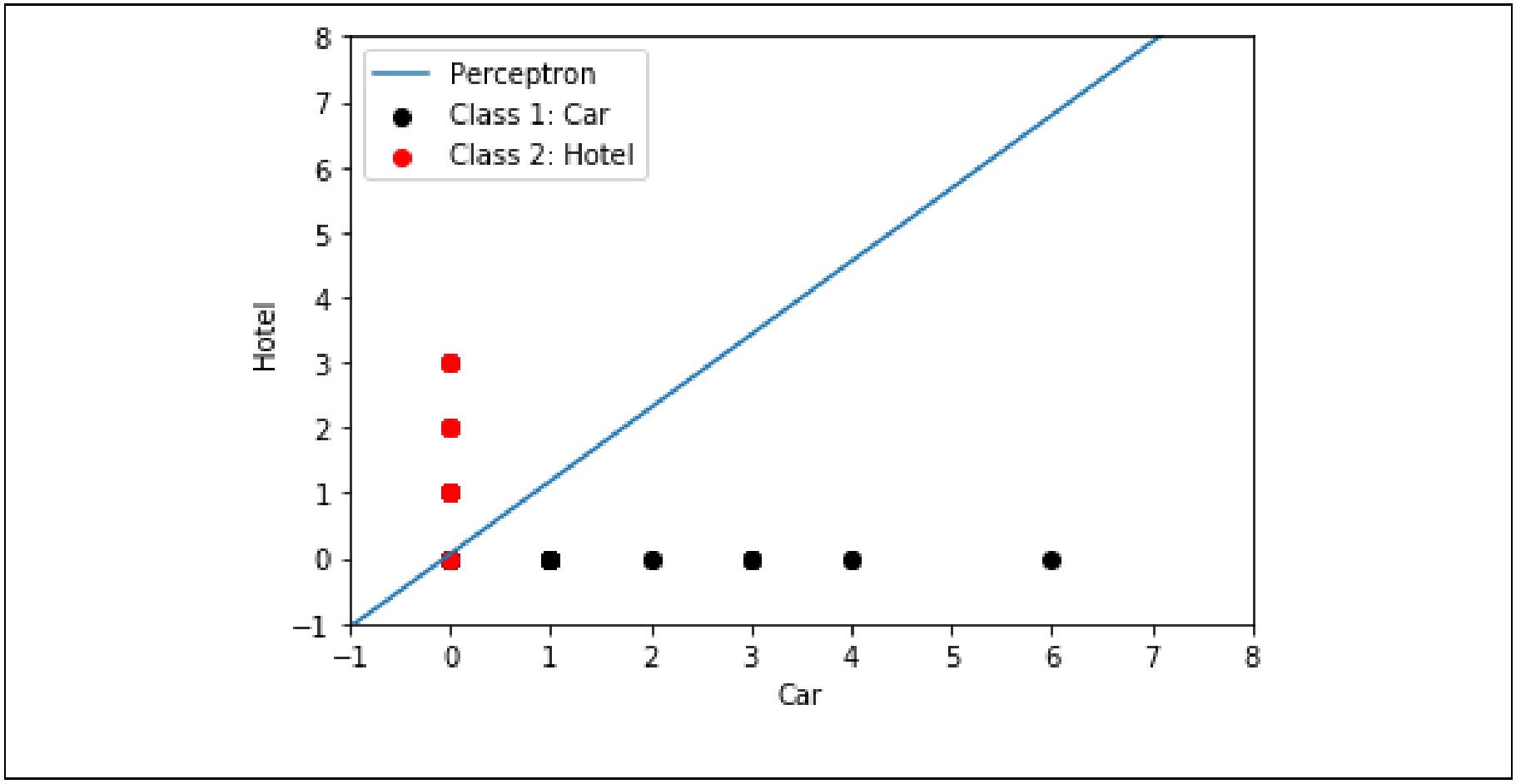
Question 3: Classifier Construction

Q3. Submit Q3.m/Q3.py (Octave/Python)

(10 marks)

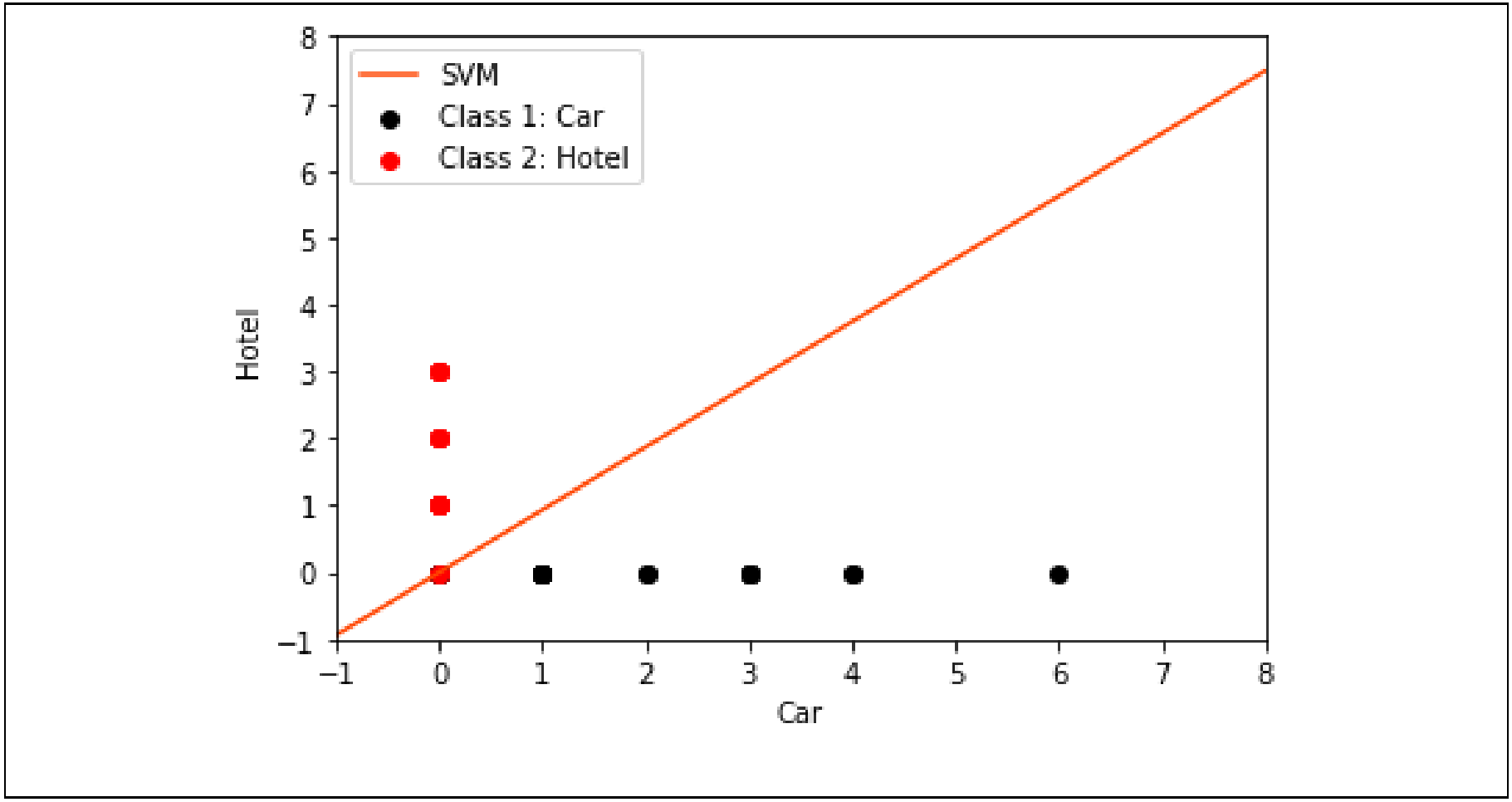
Q3(a). Plot the line obtained by perceptron onto the scatterplot using only two features: "car" and "hotel".

(10 marks)



Q2(b). Plot the line obtained by SVM onto the scatterplot using only two features: “car” and “hotel”.

(10 marks)


SVM
Class 1: Car
Class 2: Hotel
Hotel
Car
8
7
6
5
4
3
2
1
0
−1
−1
0
1
2
3
4
5
6
7
8

Question 4: Testing Stage

Q4. Submit Q4.m/Q4.py (Octave/Python)

(10 marks)

Q4(a). Write down the error obtained by perceptron and support vector machine using the two features of the testing samples.

(7 marks)

Both perceptron and svm do not converge in their training procedures, so the classifier is not obtained.

Perceptron does not converge with two features since this dataset contains an observation 0,0 [car, hotel] which belongs to both classes. This means that there is data with neither of the keywords "car" and "hotel" which belong to both the classes, and it is difficult for the perceptron, a linear classifier, to classify the data. Thus, the classifier cannot be obtained.

SVM does not converge with two features since this dataset containains an observation 0,0 [car, hotel] which belongs to both classes. This means that there is data with neither of the keywords "car" and "hotel" which belong

to both the classes, and it is difficult for the linear version of SVM to classify the data. Thus, we cannot obtain the classifier.

Q4(b). Write down the error obtained by perceptron and support vector machine using all features/variables.

(7 marks)

Perceptron: 90%

Support Vector Machine: 100%

Q4(c). Which approach gives higher accuracy? (two samples vs all). For the better approach, explain why it can obtain better accuracy.

(6 marks)

The all features / keywords approach gives a higher accuracy since some reviews contain neither the keyword "car" not "hotel". This makes the classification an impossible task for the Perceptron and SVM which are linear in nature.

To solve this problem, by including all the keywords we ensure that both the classes contain different observations and more information is available for each observation. This helps on the training procedure of both the Perceptron and SVM.