CS4 Tutorial 2

k-Nearest Neighbour Classifiers

1. Three cases from a regression system for estimating blood-alcohol content and shown in the figure below. The input features are; Gender, Framesize (i.e. weight), Amount of alcohol in units, Meal (None, Snack, Full), Duration of drinking session.

N-1	
Gender	Male
FrameSize	1
Amount	1
Meal	snack
Duration	60
BAC	0.2

N-3	
Gender	Female
FrameSize	4
Amount	4
Meal	full
Duration	90
BAC	0.8

N-55	
Gender	Male
FrameSize	1
Amount	3
Meal	snack
Duration	120
BAC	0.7

   a. Propose a similarity metric for comparing cases such as these. You can assume that the range for; Framesize is 1-6, Amount is 1-16, Duration is 20-300. The similarity metric should take account of the fact that Meal is an ordered feature.

   b. Use this metric to calculate the similarities between N-1 and N-3 and N-1 and N-55.

2 . Two cases from a CBR system for estimating the price of secondhand motorcars are shown in the figure below.

CP 007	
Manufacturer	Ford
Model	Fiesta
Age	5
Engine Size	1,000
Fuel	Petrol
Mileage	65,000
Bodywork	Excellent
Price	£3,100

CP 014	
Manufacturer	Citroen
Model	BX
Age	6
Engine Size	1,800
Fuel	Diesel
Mileage	65,000
Bodywork	Good
Price	£4,500

   (i) Propose a similarity metric that might be used in a *k*-Nearest Neighbour case retrieval system for such a case base (i.e. using exhaustive search). Each case has 7 features, 4 symbolic features and 3 numeric.

   (ii) If the Bodywork feature is an ordered feature that has the possible values {Poor, Fair, Good, Excellent} how might the similarity metric be modified to accommodate this similarity information.

   (iii) How might this retrieval system be improved using feature weights?

Q7. 2002

7. The figure below shows two cases from a Case-Based Reasoning (CBR) system for estimating the resale price for second hand cars. The objective is to take a target case for which a resale price needs to be estimated and retrieve cases from the case-base to use to predict this price.

CP 007

Manufacturer	Ford
Model	Fiesta
Age	5
Engine Size	1,000
Fuel	Petrol
Mileage	65,000
Bodywork	Excellent
Price	€3,100

CP 014

Manufacturer	Citroen
Model	BX
Age	6
Engine Size	1,800
Fuel	Diesel
Mileage	65,000
Bodywork	Good
Price	€4,500

a) Propose a nearest-neighbour metric for identifying similar cases in such a CBR system.

(7 marks)

b) Discuss how a set of say three retrieved cases might be used to determine a price for the target car.

(6 marks)

c) Show how adaptation might be used to adjust prices to account for differences between retrieved and target cases. Propose some example adaptation rules or operators.

(7 marks)