

# 3BA2 Tutorial 1

## $k$ -NN

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		N-3		N-55	
Gender	Male	Gender	Female	Gender	Male
Weight	70	Weight	60	FrameSize	75
Amount	1	Amount	4	Amount	3
Meal	snack	Meal	full	Meal	snack
Duration	60	Duration	90	Duration	120
BAC	0.2	BAC	0.8	BAC	0.7

- a. Propose a similarity metric for comparing cases such as these. You can assume that the range for; Weight is 50-150, 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		CP 014	
Manufacturer	Ford	Manufacturer	Citroen
Model	Fiesta	Model	BX
Age	5	Age	6
Engine Size	1,000	Engine Size	1,800
Fuel	Petrol	Fuel	Diesel
Mileage	65,000	Mileage	65,000
Bodywork	Excellent	Bodywork	Good
Price	£3,100	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?

3. Consider the following data set with three Boolean predictive attributes, W, X, Y and Boolean classification C.

W	X	Y	C
T	T	T	T
T	F	T	F
T	F	F	T
F	T	T	F
F	F	F	F

We now encounter a new example: W=F, X=F, Y=T.

- (a) Using 3-Nearest Neighbor classifier, which class will be assigned to the example?