Tutorial 8:
Current Best Learning, Decision Trees, Naive Bayes

Name	Student no.

For this tutorial, you can discuss the questions in small groups (up to 4 students). Individually submit the answers to each of the 3 Questions.

Question 1

This problem asks you to search a *hypothesis space*. Assume that we have four examples (see below) where attributes take values as indicated in the header. The target attribute/concept is *EnjoySport*.

	Sky (Sunny, Cloudy, Rain)	*AirTemp (Warm, Cold)*	*Humidity (Normal, High)*	*Wind (Weak, Strong)*	*Water (Warm, Cool)*	*Forecast (Same, Change)*	*EnjoySport (Yes, No)*
1	Sunny	Warm	Normal	Strong	Warm	Same	No
2	Sunny	Warm	High	Strong	Warm	Same	Yes
3	Rain	Cold	High	Strong	Warm	Change	No
4	Sunny	Warm	High	Strong	Cool	Change	Yes

A hypothesis regarding what it means to *EnjoySport* is represented as a pattern **< v1, v2, v3, v4, v5, v6 > ↔ Yes**, where **v1** can take the values *Sunny*, *Cloudy* or *Rain* (according to the first attribute *Sky*) or a wildcard value ***** (matching all possible values), **v2** can take the values of *AirTemp* plus *****, and so forth for all the attributes. All examples matching the pattern are classified as *Yes* (i.e. part of the concept *EnjoySport*), those not matching the pattern are classified as *No*. In addition there is one special pattern **< nil > ↔ Yes** which classifies all examples as *No*.

a) What size is the hypothesis space? (ie: how many possible hypotheses are there, following the rules above).

b) Rank the following hypotheses according to *EnjoySport=Yes* specificity (1-most specific, 5-least specific):

A= **< *, *, High, *, *, * > ↔ Yes.**

B= **< *, *, Normal, Weak, *, Change > ↔ Yes.**

C= **< Rain, Cold, High, Strong, Warm, Change > ↔ Yes.**

D= **< nil > ↔ Yes.**

E= **< *, Cold, Normal, Strong, Cool, Same > ↔ Yes.**

c) Using Current-best-hypothesis learning, track the currently best (most specific) hypothesis, going through examples 2-4, starting with

H1= **< *, Cold, High, *, *, * >** ↔ **Yes** (consistent with example 1)

Question 2

Ernie's entertainment park has a merry-go-round that makes some people sick. Ernie has recently collected data to help resolve which attributes cause this condition.

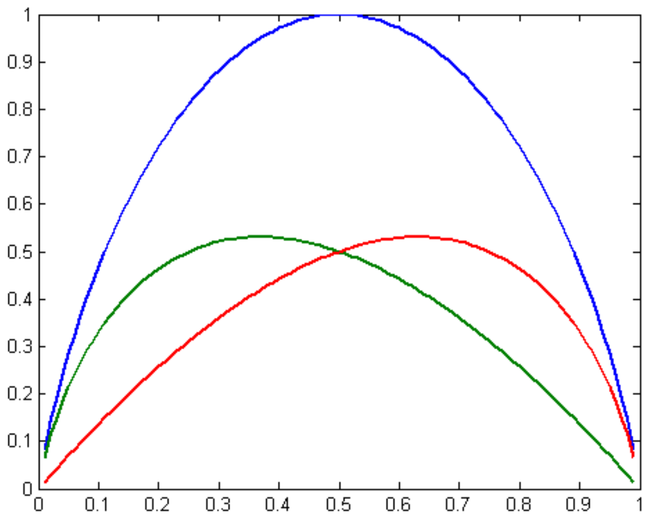
	Height	Weight	Age	Gender	Sick?
D1	Tall	Fat	Young	Female	Yes
D2	Tall	Thin	Middleage	Male	Yes
D3	Short	Medium	Old	Male	No
D4	Medium	Medium	Old	Female	No
D5	Medium	Fat	Young	Male	No
D6	Tall	Thin	Young	Female	Yes
D7	Short	Medium	Middleage	Male	No
D8	Medium	Fat	Young	Female	No
D9	Tall	Thin	Old	Female	Yes
D10	Tall	Thin	Young	Female	Yes
D11	Short	Medium	Middleage	Female	No
D12	Tall	Medium	Young	Male	Yes
D13	Tall	Fat	Young	Female	Yes
D14	Short	Thin	Old	Male	No
D15	Medium	Thin	Old	Female	Yes
D16	Tall	Fat	Young	Female	Yes
D17	Tall	Thin	Middleage	Male	Yes
D18	Short	Thin	Young	Male	No
D19	Medium	Fat	Old	Female	No
D20	Tall	Thin	Young	Male	Yes

$$Gain(A) = I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) - Remainder(A)$$

$$Remainder(A) = \sum_{i=1}^{v} \frac{p_i + n_i}{p+n} I\left(\frac{p_i}{p_i + n_i}, \frac{n_i}{p_i + n_i}\right)$$

$$I(P(v_1), \ldots, P(v_m)) = \sum_{i=1}^{m} - P(v_i) \log_2 P(v_i)$$

$$I\left(\frac{p}{p+n}, \frac{n}{p+n}\right) = -\frac{p}{p+n}\log_2\frac{p}{p+n} - \frac{n}{p+n}\log_2\frac{n}{p+n}$$



a) Which attribute has the greatest information gain with respect to the classification of motion sickness?

b) What is the greatest information gain?

c) Create a full ID3 decision tree.

d) Do you think that this decision tree would be sufficient for Ernie to warn nearly everyone at risk? Why / why not?

Question 3

Construct a Naive Bayes model for the same data as Question 2 (D1-D20) and apply it to the test data below (T1-T2), detailing the model and the test results.

What you need to know: (reviewing theory)

The Naive Bayes assumption is that the joint probability of the attributes (e.g. P(*x*new|*c*new,d) can be decomposed into the product of the individual attribute probabilities. The resulting classification rule is as follows:

$$P(c_{new} \mid x_{new}, d) \propto P(x_{new} \mid c_{new}, d) P(c_{new} \mid d) \propto P(c_{new} \mid d) \prod_{i=1}^{m} P(x_{i,new} \mid c_{new}, d)$$

*x*new is a new input vector, *c*new is a possible class (output) for this input, d is the dataset. The P(*c*new|*x*new,d) values must sum to 1 over the possible classes *c*new. You can do this renormalisation once you have calculated the un-normalised values for each of the (2) output classes.

	Height	Weight	Age	Gender	Sick?
T1	Tall	Medium	Young	Female	(Yes)
T2	Medium	Medium	Middleage	Male	(No)