AI2 Module 3
Final Notes

Jacques Fleuriot

Warning

These notes are for guidance only and do not provide an exhaustive list of what is or isn't examinable. It may very helpful to be aware of what follows though!

Example of things that you should be able to do

- provide definitions: you should be able to define terms such as supervised/unsupervised/reinforcement learning, decision tree, training set, test set, false positive, false negative etc. You may be asked to give illustrative examples as well.

- give decision tree learning formulae such as that for **entropy** I of a set of examples, that for the **remainder** after a split, and the one for the **information gain**.

- give descriptions of learning algorithms. These may be given using pseudo-code (as in the slides), for example. You should know algorithms such as decision-tree learning, current-best learning (CBL), version-space learning (candidate elimination algorithm), for example.

- describe strengths and weaknesses of algorithms e.g. the effects of noisy data on them.
  - for decision tree algorithm, for example, you should know what happens when it runs out of examples or attributes (with examples left) and explain why such a situation might arise.
  - you should know what the problems are with CBL e.g. difficulty

in finding good heuristics.

- say how current-best learning deals with false positive and false negatives and give examples e.g. when learning logical descriptions.
- define terms such as recurrent and feed-forward, activation function, threshold, linearly separable, etc. with respect to neural networks.
- give formulae for weighted input to neuron, activation functions such as the step, sign, and sigmoid, output value, weight update for perceptron, error functions such as $\frac{1}{2}\sum_e (T_e - O_e)^2$.

- you may be given an error function and be asked to compute the slope of the error surface (used for gradient descent).
- devise a network for computing a simple function e.g. a boolean function (XOR was given in lecture).
- give perceptron-learning algorithm (including weight update formula) and explain all the symbols/terms given in formulae e.g. learning rate.
- simulate the perceptron learning algorithm on one or more examples. Show the formulae that you use in your computations i.e. don't just give sequences of numbers and/or the final result.

This ensures that you get as many marks as possible if you make a mistake in your derivation.

- give the gradient descent algorithm and know what its weaknesses are e.g. problems with local minima etc.
- give the back-propagation-learning algorithm.
- give the weight update rule used in back-propagation and know what the Δ error terms are for the output and hidden units.
- simulate the back-propagation learning algorithm on one or more examples. Same remarks as for perceptron learning apply.

- define terms such as qualification problem, monotonic, non-monotonic, etc.
- mention some of the problems with default logic.
- define probability terms such as elementary event, sample space, event, joint distribution, marginal distribution.
- give the axioms of probability.
- define conditional probability rule, Bayes' theorem (rule). You need to know these as they are central to Bayesian inference. Do not confuse them with one another!

- know how to compute a marginal distribution from a given distribution. You must be able to write down the formula for doing this (i.e. the "summing out" formula) and then manipulate it.
- define what it means for two events to be independent i.e. give the formulae for conditional or standard independence.
- derive formula such as:

$$\mathtt{Pr}\{A = a, B = b | C = c\} = \mathtt{Pr}\{A = a | B = b, C = c\}\mathtt{Pr}\{B = b | C = c\}$$

(see slide 3-9 for this one) This particular result is useful to know in any case!

- give the algorithm for constructing a Bayesian network.
- explain why "causal" structure is preferable for a Bayesian network.
- make inference using a Bayesian network i.e. compute particular probabilities and/or probability distributions.
- use Bayes' theorem to change from a causal to a diagnostic inference and vice-versa.
- give the algorithm for "inference by simulation" and explain the idea behind likelihood weighting.

- define notion of decision network and associated concepts such utility, expected utility, maximum expected utility, value of information etc. You need to know the various formulae.
- You may be asked to compute the expected utility of a particular action or you may be given utility formulae and appropriate probabilities, and then be asked to find what the best action is for a rational agent.

Don't forget

- to write and manipulate the formulae (e.g. probability ones or weight updates) that you are dealing with for as long as possible and substitute actual values as the last step.
- when giving algorithms, to show the initialization, main body, and termination condition(s) explicitly especially if you're not using pseudo-code.
- that you need to practice your probability computations (as well as other types of computations!).
- use Russell and Norvig, as well, when revising.

You won't be asked

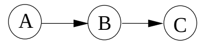
1. to write Prolog code.
2. to perform lengthy numerical calculations but you may be asked to follow simple ones and may need to develop formulas required in such a computation.
3. to **derive** the formulae for the error terms Δ_i at the hidden and output units. Remember: you need to know what they are though.
4. about Dynamic Belief Network

Finally

- do past exam papers as they provide an indication of the level/type of questions to be expected and they have model answers. They are available online from Informatics teaching web pages.
- I hope that you enjoyed the course.

September 2002 MCQ

Given the following Bayesian network with binary variables A, B, and C,



a formula for $\texttt{Pr}\{A = a|B = b\}$ in terms of probabilities from the conditional probability tables (CPTs) attached to the nodes of the network is given by:

1. $\texttt{Pr}\{A = a|B = b\} = \texttt{Pr}\{A = a\}$
2. $\texttt{Pr}\{A = a|B = b\} = \frac{\texttt{Pr}\{A=a,B=b\}}{\texttt{Pr}\{B=b\}}$
3. $\texttt{Pr}\{A = a|B = b\} = \texttt{Pr}\{B = b\}$
4. $\texttt{Pr}\{A = a|B = b\} = \frac{\texttt{Pr}\{B=b|A=a\}\texttt{Pr}\{A=a\}}{\sum_v \texttt{Pr}\{B=b|A=v\}\texttt{Pr}\{A=v\}}$
5. $\texttt{Pr}\{A = a|B = b\} = \frac{\texttt{Pr}\{B=b|A=a\}\texttt{Pr}\{A=a\}}{\texttt{Pr}\{B=b\}}$