

3BA2 Tutorial

Supervised Learning

1. 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) If we apply the Naive Bayes method without the 0-probability correction, what probability is assigned to the two possible values of C?
(8 marks)
- (b) What Information Gain measures would be assigned to the features by a decision tree learner? Which feature will be selected at the root of the tree? How will the example be classified?
(8 marks)
- (c) Using 3-Nearest Neighbour classifier, which class will be assigned to the example?
(4 marks)
2. (a) Explain the distinction between Lazy and Eager techniques in Machine Learning.
(6 marks)
- (b) Explain why a single layer feedforward network can only learn patterns that are linearly separable.
(6 marks)
- (c) Explain why training a multi-layer feedforward neural network is considerably more difficult than training a single layer network.
(8 marks)

3BA2 ML Exam Question '03

4. (a) Describe the architecture of a single layer neural network that can learn to associate an output vector \mathbf{t} with a general input vector \mathbf{s} , given that it can be trained with a training set of pairs $\mathbf{s} : \mathbf{t}$.
- (b) Describe an algorithm for training this network using the training set of pairs $\mathbf{s} : \mathbf{t}$. You may assume, in order to simplify the training, that the data is in bipolar format.
- (c) Explain why this architecture can only learn classification problems where the classes are linearly separable.

6. Consider the following Student Exam Performance Data taken from "The Essence of Artificial Intelligence", A. Cawsey, Prentice Hall Europe, 1998:

No	Student	First last year?	Male?	Works hard?	Drinks?	First this year?
1	Richard	+	+	-	+	+
2	Alan	+	+	+	-	+
3	Alison	-	-	+	-	+
4	Jeff	-	+	-	+	-
5	Gail	+	-	+	+	+
6	Simon	-	+	+	+	-

Useful facts: $-(1/3)\log_2(1/3)\approx.528$; $-(2/3)\log_2(2/3)\approx.390$; $-(1/2)\log_2(1/2)=.5$;
 $-(1/4)\log_2(1/4)=.5$; $-(3/4)\log_2(3/4)\approx.311$

- (a) What is the entropy of this data set with respect to the target function classification ("first this year")? (2 marks)
- (b) Construct the decision tree that would be built with Information Gain for this data set. Show your work for selection of the root feature in your tree. Do not take feature Student into account. (6 marks)
- (c) Using your decision tree from (b), how would you classify the following instance:

No	Student	First last year?	Male?	Works hard?	Drinks?	First this year?
7	Matthew	-	+	-	+	??

(2 marks)

- (d) Use Naïve Bayes to give the likelihood for "first this year?" for the given instance. Then indicate what prediction Naïve Bayes would make. Provide the probability table for each of the predicting features. (6 marks)
- (e) A good measure of distance between two instances with categorical features is the number of features which have different values (the overlap metric). Using this measure of distance what classification ("+" or "-") would be predicted by using 3-NN for the instance in (c)?

(4 marks)