

## CSc 74010 Fall 2011, Homework 9

- The following table gives some examples of recent book selections I have made on the Orinoco website (Orinoco is the world's least well known online bookstore).

Examples	Attributes					Will Buy
	New	Paper	Know	Lang	Type	
$X_1$	N	N	Y	Eng	Thriller	Y
$X_2$	N	N	Y	Sp	Romance	N
$X_3$	Y	N	N	Eng	Detective	Y
$X_4$	N	Y	Y	Sp	Romance	Y
$X_5$	N	Y	N	Sp	Thriller	N
$X_6$	Y	N	Y	Eng	Literature	Y
$X_7$	Y	N	N	Fr	Detective	N
$X_8$	N	N	Y	Eng	Romance	Y
$X_9$	Y	Y	N	Sp	Detective	N
$X_{10}$	Y	Y	Y	Sp	Literature	N
$X_{11}$	N	N	N	Fr	Romance	N
$X_{12}$	Y	Y	Y	Sp	Detective	Y

This records whether or not the book is *Newly* published, is a *Paperback* or not, whether I *Know* the author (that is whether I have previously bought a book by the same author), what *Language* the book was originally written in (English, French or Spanish), and what genre the book is from (Thriller, Romance, Detective, or Literature). The site also records whether or not I actually bought the book (or just browsed it).

Use the decision tree learning algorithm from the notes to construct a decision tree that Orinoco could use to predict whether I am likely to want to purchase any new books that they start to stock.

As ever, you should explain how the algorithm builds the decision tree, not just give the solution.

(50 points)

- How might Orinoco use linear regression instead of a decision tree to decide if I will be interested in a particular new book?

(20 points)

- What property of linear regression might make it problematic to use linear regression in place of a decision tree?

(10 points)

- It would be possible for Orinoco to use a genetic algorithm to construct the decision tree rather than the decision tree learning algorithm that we covered in class. Explain the process you would use to apply a genetic algorithm in this way.

(20 points)