CS 188 Section Handout: Classification

1 Introduction

In this set of exercises we will experiment with a binary classification problem. The data comprise a training set of *feature* vectors with corresponding *class* labels, and a test set of unlabeled feature vectors which we will attempt to classify with a decision tree, a naive Bayes model, and a perceptron.

Scenario: You are a geek who hates sports. Trying to look cool at a party, you join a lively discussion on professional football and basketball. You have no idea who plays what, but fortunately you have brought your CS188 notes along, and will build some classifiers to determine which sport is being discussed.

Training data: Somehow you come across a pamphlet from the Atlantic Coast Conference Basketball Hall of Fame, as well as an Oakland Raiders team roster. You study these to create the following table:

Sport	*Position*	*Name*	*Height*	*Weight*	*Age*	*College*
Basketball	Guard	Michael Jordan	6’06”	195	43	North Carolina
Basketball	Guard	Vince Carter	6’06”	215	29	North Carolina
Basketball	Guard	Muggsy Bogues	5’03”	135	41	Wake Forest
Basketball	Center	Tim Duncan	6’11”	260	29	Wake Forest
Football	Center	Vince Carter	6’02”	295	23	Oklahoma
Football	Kicker	Tim Duncan	6’00”	215	27	Oklahoma
Football	Kicker	Sebastian Janikowski	6’02”	250	27	Florida State
Football	Guard	Langston Walker	6’08”	345	27	California

Test data: You wish to determine which sport is played by these two subjects of discussion:

Sport	*Position*	*Name*	*Height*	*Weight*	*Age*	*College*
?	Guard	Charlie Ward	6’02”	185	35	Florida State
?	Defensive End	Julius Peppers	6’07”	283	26	North Carolina

2 Decision Trees

A decision tree classifies a feature vector by asking "questions" of its *attributes*, and returns a predicted "decision" based on a structure learned to represent the training data.

Entropy: What is the information content of the initial training distribution of *Sport*?

Information gain: Calculate the information gain for the decision stumps created by first splitting on $Position$, $Name$, or $College$? Which of these compact decision trees perfectly classifies the training data?

Feature Engineering: Decision trees can represent any function of discrete attribute variables. For continuous variables ($Height$, $Weight$, and Age), show how to reasonably cast these as Boolean variables.

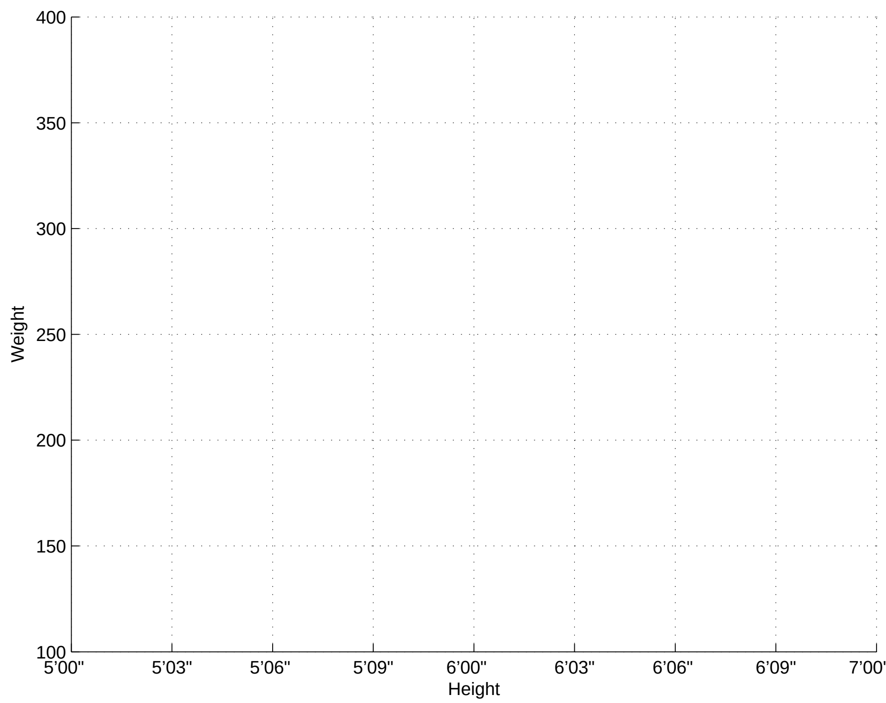
Expressiveness: Draw a few decision trees that each correctly classify the training data, and show how their predictions vary on the test set.

3 Perceptrons

A perceptron can be applied to input features with continuous values. The weights of the perceptron define hyperplanes that represent decision boundaries for classification. Assume we are only using the input features $f_1 = Height$ and $f_2 = Weight$, along with a bias $f_0 = 1$.

Structure: Draw the networks for a binary and a multi-class perceptron sports classifier.

Linear separability: A binary or two-class perceptron can only represent linearly separable functions. Plot the training data below to determine if there exists a linear separator.



Learning: Without using the learning algorithm for perceptron weight updates, can you use the plot above to guess values for the weights of a binary perceptron that correctly classifies all the training data?

4 Naive Bayes Model

In a naive Bayes model, features are conditionally independent given the generating class.

Structure: Draw a graphical model for the naive Bayes sports classifier.

Estimation: Compute empirical distributions with the maximum-likelihood estimate:

$\hat{P}(S)$

B	
F	

$\hat{P}(P \mid S)$

G	B	
C	B	
G	F	
C	F	
K	F	

$\hat{P}(N \mid S)$

MJ	B	
VC	B	
TD	B	
MB	B	
VC	F	
TD	F	
LW	F	
SJ	F	

$\hat{P}(H > 6\text{'}03\text{"} \mid S)$

$True$	B	
$False$	B	
$True$	F	
$False$	F	

$\hat{P}(W > 200 \mid S)$

$True$	B	
$False$	B	
$True$	F	
$False$	F	

$\hat{P}(A > 28 \mid S)$

$True$	B	
$False$	B	
$True$	F	
$False$	F	

$\hat{P}(C \mid S)$

NC	B	
WF	B	
O	F	
FS	F	
C	F	

Smoothing: Apply *add-one* (Laplace) smoothing to distributions, allow possible unknown (UNK) values:

$\hat{P}(S)$

B	
F	

$\hat{P}(P \mid S)$

G	B	
C	B	
UNK	B	
G	F	
C	F	
K	F	
UNK	F	

$\hat{P}(N \mid S)$

MJ	B	
VC	B	
TD	B	
MB	B	
UNK	B	
VC	F	
TD	F	
LW	F	
SJ	F	
UNK	F	

$\hat{P}(H > 6\text{'}03\text{"} \mid S)$

$True$	B	
$False$	B	
$True$	F	
$False$	F	

$\hat{P}(W > 200 \mid S)$

$True$	B	
$False$	B	
$True$	F	
$False$	F	

$\hat{P}(A > 28 \mid S)$

$True$	B	
$False$	B	
$True$	F	
$False$	F	

$\hat{P}(C \mid S)$

NC	B	
WF	B	
UNK	B	
O	F	
FS	F	
C	F	
UNK	F	

Inference: For the first test example, compute the posterior probability of *Sport* using the smoothed and unsmoothed empirical probability estimates.

Extra: Charlie Ward and Julius Peppers were athletes who each competed in both football *and* basketball during college. If you trained these classifiers correctly, you can predict what professional sports they chose.