CS 188 Spring 2011	Introduction to Artificial Intelligence	Section 11 Classification

Pacman and Mrs. Pacman have been searching for each other in the Maze. Ms. Pacman has been pregnant with a baby, and just this morning she has given birth to Pacbaby. Ms. Pacman wants to teach Pacbaby to recognize his father, using a set of Polaroids of Pacman. She also has several pictures of ghosts to use as negative examples. Because the polaroids are black and white, and were taken from strange angles, Ms. Pacman has decided to teach Pacbaby to identify Pacman based on more salient features: the presence of a bowtie (b), hat (h), or mustache (m).

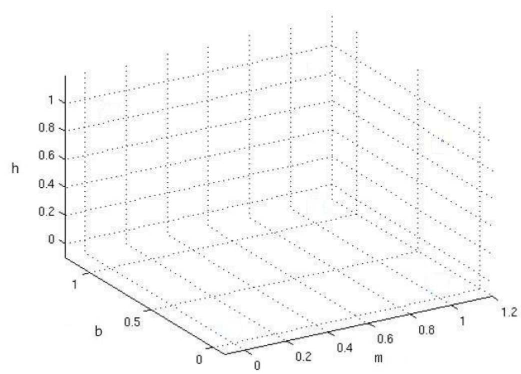
The following table summarizes the content of the Polaroids. Each binary feature is represented as 1 (meaning the feature is present) or 0 (meaning it is absent). The subject y of the photo is encoded as +1 for Pacman or -1 for ghost.

(m)	(b)	(h)	Subject (y)
0	0	0	+1
1	0	0	+1
1	1	0	+1
0	1	1	+1
1	0	1	-1
1	1	1	-1

Q1. Pacbaby Classifiers

Suppose Pacbaby has a perceptron based brain, meaning his "internal classifier" learns through perceptron updates and is limited to learning linear classification rules. Further, suppose Pacbaby "augments" each example it sees with a bias feature that is always equal to 1 (This allows Pacbaby to learn decision rules with boundaries that do not pass through the origin)

(a) Will Pacbaby be able to learn a rule that makes no mistakes on the set of Polaroids? In other words, is the training set linearly separable? (Plot the training data)



(b) Suppose there was another Polaroid of a character without a mustache or a hat, but who was wearing a bowtie. If this Polaroid was of Pacman, would the data be linearly separable? What if it contained a ghost?

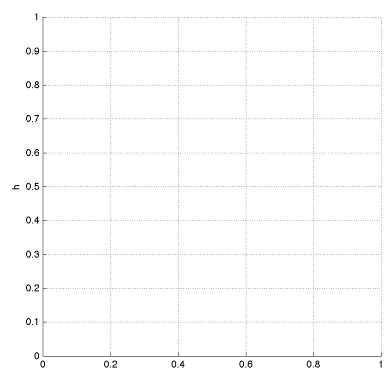
(c) Suppose we start with the training weights $[-1, 1, -1, -1]$, and wish to train a perceptron on the above data. Perform two updates of the Perceptron algorithm, processing the training data in the order they appear. The first 3 weights correspond to the features (m),(b),(h), respectively. The last weight corresponds to the bias feature. If a training example has weight exactly 0, classify it as +1.

	w_1	w_2	w_3	w_4
Initial weights	-1	1	-1	-1
Training: (0,0,0,1)→ +1				
Training: (1,0,0,1)→ +1				

Now let us assume that Pacbaby has a K-nearest-neighbors brain. We will use a new set of polaroids. Pacbaby's vision is blurry, so features are no longer binary: a feature may be anywhere between 1 (the feature is definitely present) or 0 (the feature is definitely absent). Moustaches have been banned in Pac-land, so our only features are bowties and beards. The subject y of the photo is encoded as +1 for Pacman or -1 for ghost.

(b)	(h)	Subject (y)
0	0	+1
0.5	0	+1
1	1	+1
0.5	1	+1
0.5	0.5	-1
1	0	-1
0	1	-1

(d) Plot the training data. Is this data linearly separable? Does that matter?



(e) Use K-nearest-neighbors, with Manhattan distance as the metric, and with $k = 3$ and $k = 5$ as the number of nearest neighbors, to classify the following polaroids.

(b)	(h)	$k = 3$	$k = 5$
0.2	0.2		
0.2	0.8		
0.8	0.8		
1	1		
1	0		