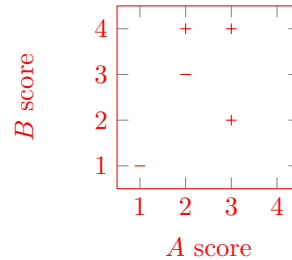


1 Perceptron

You want to predict if movies will be profitable based on their screenplays. You hire two critics A and B to read a script you have and rate it on a scale of 1 to 4. The critics are not perfect; here are five data points including the critics' scores and the performance of the movie:

#	Movie Name	A	B	Profit?
1	Pellet Power	1	1	-
2	Ghosts!	3	2	+
3	Pac is Bac	2	4	+
4	Not a Pizza	3	4	+
5	Endless Maze	2	3	-



- (a) First, you would like to examine the linear separability of the data. Plot the data on the 2D plane above; label profitable movies with + and non-profitable movies with - and determine if the data are linearly separable.

The data are linearly separable.

- (b) Now you decide to use a perceptron to classify your data. Suppose you directly use the scores given above as features, together with a bias feature. That is $f_0 = 1$, $f_1 =$ score given by A and $f_2 =$ score given by B.

Run one pass through the data with the perceptron algorithm, filling out the table below. Go through the data points in order, e.g. using data point #1 at step 1.

step	Weights	Score	Correct?
1	$[-1, 0, 0]$	$-1 \cdot 1 + 0 \cdot 1 + 0 \cdot 1 = -1$	yes
2	$[-1, 0, 0]$	$-1 \cdot 1 + 0 \cdot 3 + 0 \cdot 2 = -1$	no
3	$[0, 3, 2]$	$0 \cdot 1 + 3 \cdot 2 + 2 \cdot 4 = 14$	yes
4	$[0, 3, 2]$	$0 \cdot 1 + 3 \cdot 3 + 2 \cdot 4 = 17$	yes
5	$[0, 3, 2]$	$0 \cdot 1 + 3 \cdot 2 + 2 \cdot 3 = 12$	no

Final weights: $[-1, 1, -1]$

- (c) Have weights been learned that separate the data? With the current weights, points will be classified as positive if $-1 \cdot 1 + 1 \cdot A + -1 \cdot B \geq 0$, or $A - B \geq 1$. So we will have incorrect predictions for data points 3:

$$-1 \cdot 1 + 1 \cdot 2 + -1 \cdot 4 = -3 < 0$$

and 4:

$$-1 \cdot 1 + 1 \cdot 3 + -1 \cdot 4 = -2 < 0$$

Note that although point 2 has $w \cdot f = 0$, it will be classified as positive (since we classify as positive if $w \cdot f \geq 0$).

- (d) More generally, irrespective of the training data, you want to know if your features are powerful enough to allow you to handle a range of scenarios. Circle the scenarios for which a perceptron using the features above can indeed perfectly classify movies which are profitable according to the given rules:

- (a) Your reviewers are awesome: if the total of their scores is more than 8, then the movie will definitely be profitable, and otherwise it won't be. **Can classify (consider weights $[-8, 1, 1]$)**
- (b) Your reviewers are art critics. Your movie will be profitable if and only if each reviewer gives either a score of 2 or a score of 3. **Cannot classify**
- (c) Your reviewers have weird but different tastes. Your movie will be profitable if and only if both reviewers agree. **Cannot classify**

2 Maximum Likelihood

A Geometric distribution is a probability distribution of the number X of Bernoulli trials needed to get one success. It depends on a parameter p , which is the probability of success for each individual Bernoulli trial. Think of it as the number of times you must flip a coin before flipping heads. The probability is given as follows:

$$P(X = k) = p(1 - p)^{k-1} \tag{1}$$

p is the parameter we wish to estimate.

We observe the following samples from a Geometric distribution: $x_1 = 5, x_2 = 8, x_3 = 3, x_4 = 5, x_5 = 7$. What is the maximum likelihood estimate for p ?

$$L(p) = P(X = x_1)P(X = x_2)P(X = x_3)P(X = x_4)P(X = x_5) \tag{2}$$

$$= P(X = 5)P(X = 8)P(X = 3)P(X = 5)P(X = 7) \tag{3}$$

$$= p^5(1 - p)^{23} \tag{4}$$

$$\log(L(p)) = 5 \log(p) + 23 \log(1 - p) \tag{5}$$

$$\tag{6}$$

We must maximize the log-likelihood of p , so we will take the derivative, and set it to 0.

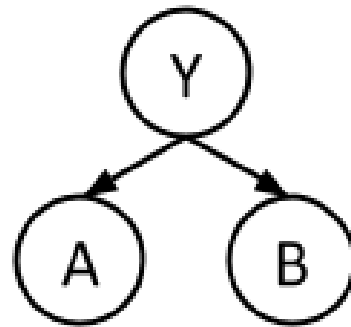
$$0 = \frac{5}{p} - \frac{23}{1 - p} \tag{7}$$

$$p = 5/28 \tag{8}$$

3 Naive Bayes

In this question, we will train a Naive Bayes classifier to predict class labels Y as a function of input features A and B . $Y, A,$ and B are all binary variables, with domains 0 and 1. We are given 10 training points from which we will estimate our distribution.

A	1	1	1	1	0	1	0	1	1	1
B	1	0	0	1	1	1	1	0	1	1
Y	1	1	0	0	0	1	1	0	0	0



1. What are the maximum likelihood estimates for the tables $P(Y), P(A|Y),$ and $P(B|Y)$?

Y	$P(Y)$	A	Y	$P(A Y)$	B	Y	$P(B Y)$
0	$3/5$	0	0	$1/6$	0	0	$1/3$
1	$2/5$	1	0	$5/6$	1	0	$2/3$
		0	1	$1/4$	0	1	$1/4$
		1	1	$3/4$	1	1	$3/4$

2. Consider a new data point $(A = 1, B = 1)$. What label would this classifier assign to this sample?

$$P(Y = 0, A = 1, B = 1) = P(Y = 0)P(A = 1|Y = 0)P(B = 1|Y = 0) \quad (9)$$

$$= (3/5)(5/6)(2/3) \quad (10)$$

$$= 1/3 \quad (11)$$

$$P(Y = 1, A = 1, B = 1) = P(Y = 1)P(A = 1|Y = 1)P(B = 1|Y = 1) \quad (12)$$

$$= (2/5)(3/4)(3/4) \quad (13)$$

$$= 9/40 \quad (14)$$

$$(15)$$

Our classifier will predict label 0.

3. Let's use Laplace Smoothing to smooth out our distribution. Compute the new distribution for $P(A|Y)$ given Laplace Smoothing with $k = 2$.

A	Y	$P(A Y)$
0	0	$3/10$
1	0	$7/10$
0	1	$3/8$
1	1	$5/8$