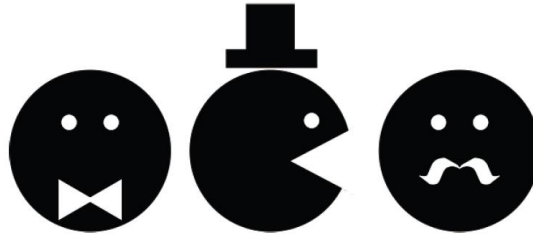


CS188 Spring 2014 Section 10: ML

Pacman and Mrs. Pacman have been searching for each other in the Maze. Mrs. Pacman has been pregnant with a baby, and just this morning she has given birth to Pacbaby (Congratulations, Pacmans!).

Because Pacbaby was born before Pacman and Mrs. Pacman reunited in the maze, he has never met his father. Naturally, Mrs. 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, Mrs. 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 |

1 Naive Pacbaby

Suppose Pacbaby has a Naive Bayes based brain.

- Write the Naive Bayes classification rule for this problem (i.e. write a formula which given a data point $x = (m, b, h)$ returns the most likely subject y). Write the formula in terms of conditional and prior probabilities. Be explicit about which parameters are involved, but you do not need to estimate them yet.

$$y^* = \arg \max_y \{p(y)p(m|y)p(b|y)p(h|y)\}$$

The parameters are the prior / conditional probabilities of seeing the feature given the class.

- Assuming no smoothing, give estimates for the parameters of the classification rule based on the Polaroids.

| | $y = +1$ | $y = -1$ |
|--------------|---------------|---------------|
| $P(y)$ | $\frac{4}{6}$ | $\frac{2}{6}$ |
| P | | |
| $P(m = 1 y)$ | $\frac{2}{4}$ | $\frac{2}{2}$ |
| $P(b = 1 y)$ | $\frac{2}{4}$ | $\frac{1}{2}$ |
| $P(h = 1 y)$ | $\frac{1}{4}$ | $\frac{2}{2}$ |

- c) Suppose a character comes by wearing a hat but without a mustache or bowtie. What would happen if Pacbaby had to guess the identity of the character?

If $y = 1$, the probability of seeing this data point $x = (0, 0, 1)$ is

$$P(y = 1)P(m = 0|y = 1)P(b = 0|y = 1)P(h = 1|y = 1) = (4/6)(2/4)(2/4)(1/4) = 1/24$$

If $y = -1$, the probability $P(m = 0|y = -1) = 0$, so the probability of seeing this is 0.

Pacbaby would classify this as Pacman.

- d) Suppose now that Pacbaby performs Laplace smoothing with strength $k = 1$ (on both the prior and class-conditional parameters). Re-estimate the parameters. Now how will Pacbaby classify this new character with the hat and without a mustache or bowtie?

| | | |
|--------------|---------------|---------------|
| | $y = +1$ | $y = -1$ |
| $P(y)$ | $\frac{5}{8}$ | $\frac{3}{8}$ |
| P | $y = +1$ | $y = -1$ |
| $P(m = 1 y)$ | $\frac{3}{6}$ | $\frac{3}{4}$ |
| $P(b = 1 y)$ | $\frac{3}{6}$ | $\frac{2}{4}$ |
| $P(h = 1 y)$ | $\frac{1}{6}$ | $\frac{1}{4}$ |

If $y = 1$, the probability of seeing this data point $x = (0, 0, 1)$ is

$$P(y = 1)P(m = 0|y = 1)P(b = 0|y = 1)P(h = 1|y = 1) = (5/8)(3/6)(3/6)(2/6) = 5/96$$

If $y = -1$, the probability is

$$P(y = -1)P(m = 0|y = -1)P(b = 0|y = -1)P(h = 1|y = -1) = (3/8)(1/4)(2/4)(3/4) = 9/256$$

Pacbaby would still classify this as Pacman.

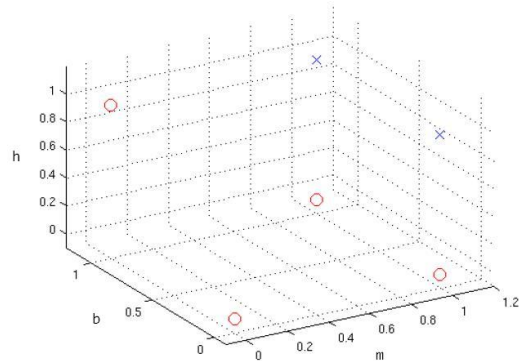
2 Perceptron Pacbaby

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).

| (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 |

- 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)

Yes, as clear in the above graph, the training set is linearly separable. Above, the blue Xs represent the ghost (-1) examples, and the red circles represent the Pacman (+1) examples.



- 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?

The data would be linearly separable if the Polaroid was of Pacman, but NOT if the Polaroid was of a ghost. If the Polaroid was of a ghost, consider the 2d plane formed by the vertices $(0,1,0)$, $(0,1,1)$, $(1,0,1)$, $(1,0,0)$. In this plane, two opposite corners of the rectangle are $+1$ and the other opposite corners are -1 ; it is not possible to separate these 2 classes in this 2d plane.

- 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: $(0,0,0,1) \rightarrow +1$ | -1 | 1 | -1 | 0 |
| Training: $(1,0,0,1) \rightarrow +1$ | 0 | 1 | -1 | 1 |

The perceptron makes a mistake on the first training example: $w^\top x = -1 < 0 \rightarrow -1$, so we add the first training example to get $[-1, 1, -1, 0]$. On the second training example it makes another mistake $w^\top x = -1 < 0 \rightarrow -1$, so we add the second training example to get weights $[0, 1, -1, 1]$.