

1 Learning (13/12/2018)

1.1 Objectives

At the end of this repetition you should be able to:

- Apply Maximum Likelihood Estimation (MLE) to estimate the parameters of a model from data.

1.2 Exercises

a PacBaby ($\approx 20min$ ¹)

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

Table 1: Matrix of Pacman polaroids data.

Suppose Pacbaby has a Naive Bayes based brain.

1. 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.
2. What are the parameters of this model? Give estimates of these parameters for the classification rule based on the Polaroids.
3. 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?

b Predict your grade ($\approx 30min$)

The Bayesian network shown in Figure 1 represents how the final mark of this class is computed. In this model, X_1, X_2, X_3 respectively denote the marks obtain by a student at the homework, the project and at the exam. The teaching assistant who marks the exam also marks the homework, which implies a slight bias in the exam correction, such that we can write $X_3 = a_1X_2 + \mathcal{N}(\mu_3, \sigma_3)$. The random variable $C = a_2X_1 + a_3X_2 + \mathcal{N}(\mu_C, \sigma_C)$ is a linear combination of the homework and project grades plus a computation error which is normally distributed. Finally, $Y = a_4C + a_5X_3 + \mathcal{N}(\mu_y, \sigma_y)$ stands for the final grade obtained by the student, which is a linear combination of the exam grades and the grades obtained for the work done during the semester plus some Gaussian noise on it due to rounding errors. Answer the following questions about this model.

1. Assume the parameters of the model are known, give the prediction rule of Y given X_1, X_2, X_3 .
2. Simplify this rule and give the minimal set of parameters of this model that are required to predict Y given X_1, X_2, X_3 .
3. Can you recognise a model seen in class?

¹Source: http://ai.berkeley.edu/sections/section_10_sp14.pdf

4. From a data set of points $d = \{(x_{1,1}, x_{2,1}, x_{3,1}, y_1), \dots, (x_{1,N}, x_{2,N}, x_{3,N}, y_N)\}$ explain how you would compute the parameters of the complete model.
5. Do the same for the minimal subset of parameters required to predict Y from X_1, X_2, X_3 .

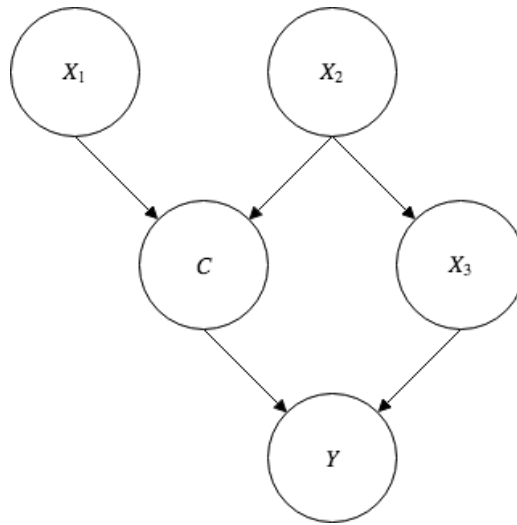


Figure 1: Bayesian network of the final grade calculation.