

INFO8006 Introduction to Artificial Intelligence

Exercises 7: Learning

Learning outcomes

At the end of this exercise session, you should be able to:

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

Exercise 1: PacBaby 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?

Exercise 2: Predict your grade

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 $X_3 = a_1 X_2 + \mathcal{N}(\mu_3, \sigma_3^2)$. The random variable $C = a_2 X_1 + a_3 X_2 + \mathcal{N}(\mu_C, \sigma_C^2)$ is a linear combination of the homework and project grades plus a computation error which is normally distributed. Finally, $Y = a_4 C + a_5 X_3 + \mathcal{N}(\mu_y, \sigma_y^2)$ 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?
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 simplest model.

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

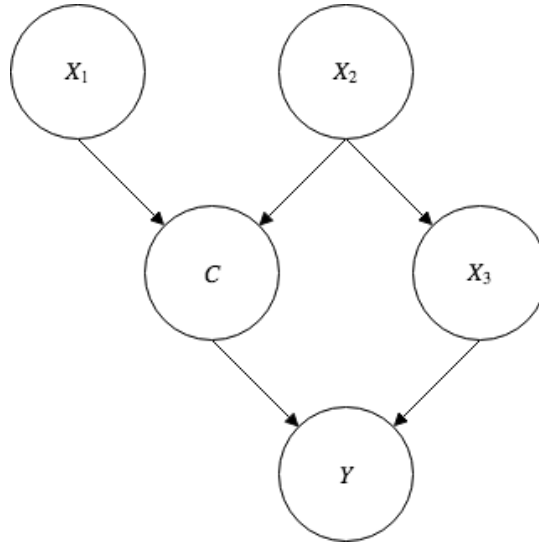


Figure 1: Bayesian network of the final grade calculation.

★ Exercise 3: Learning to play Pacman (August 2020)

You observe a Grandmaster agent playing Pacman. How can you use the moves you observe to train your own agent?

- Describe formally the data you would collect, the inference problem you would consider, and how you would solve it.
- How would you design a neural network to control your agent? Define mathematically the neural network architecture, its inputs, its outputs, its parameters, as well as the loss you would use to train it.
- Discuss the expected performance of the resulting agent when (i) the Grandmaster agent is optimal, and (ii) the Grandmaster agent is suboptimal.

★ Exercise 4: Heteroscedastic linear regression

What becomes the expression of the weight matrix W in the solution of 2.4 if the noise is different for each sample? In particular, $y_i \sim \mathcal{N}(y|W^T X, \sigma_i^2)$. Suppose you know the values σ_i . Bonus: Can you generalise this expression when the noise is correlated?

★ Exercise 5: Ridge regression

Show that linear regression with $L1$ penalisation corresponds to assuming that noise follows a Laplace distribution in contrast to the correspondence between $L2$ penalisation (least squares) and Normal noise distribution.