

Learning Probabilistic Models

Renato Lui Geh
Supervisor: Denis Deratani Mauá

1 Introduction

In this report we cite a few probabilistic models and how to perform learning on them. This report will be divided into five sections:

1. Inference on Bayesian networks
2. Bayesian learning
3. Maximum-likelihood parameter learning
4. Naive Bayes
5. Bayesian parameter learning

On (1) we will cover the definition of inference on Bayesian networks and how to perform simple queries. We will cover exact inference (which is in most cases intractable) and cite sources for approximate inference.

Next, on (2) we will take a look at the most simple of learning – Bayesian learning. We will show how to perform a simple query using the candy example presented on *Artificial Intelligence: A Modern Approach* [8] by Russel and Norvig. We will use this example throughout most of these learning methods.

We will then present Maximum-likelihood parameter learning (3), a better form of learning that is widely used in many applications. However this method performs poorly on small training sets.

Fourthly we will present Naive Bayes, a well known model that uses a simple Bayesian network composed of a single parent node and n children nodes that are d-separated (i.e. independent) of each other given their parent.

Finally, we will show how to perform Bayesian parameter learning. In this method we learn the parameters of a compiled arithmetic circuit that represents a Bayesian network.

This report is heavily influenced by the works of Russel and Norvig on AIMA [8]. It is based on the readings of chapter 14 (Probabilistic Reasoning) and 20 (Learning Probabilistic Models) of the 3rd edition.

We consider the reader has some knowledge of basic probability and a basic understanding of Bayesian networks. More specifically we consider the reader has read the following previous reports: [2, 3, 4, 5, 6, 7].

1.1 Notation

For this and all following reports we will use the notations described in this subsection.

Any function that is of the form $P(\cdot)$ is a probability function. \mathbf{X} denotes a set of variables with instances X_1, \dots, X_k . X denotes a random variable. \mathbf{E} is the set of evidence variables

and \mathbf{e} is an observed event. A prior probability is a probability function of the form $P(\mathbf{X})$. A posterior probability is of the form $P(\mathbf{X}|\mathbf{Y})$. A probability function $\mathbf{P}(\cdot)$ returns a distribution of probability.

2 Inference on Bayesian networks

Inference in any probabilistic model is to compute a posterior probability distribution of a query of variables given an observed evidence. Let \mathbf{Y} be a set of nonevidence nonquery variables Y_1, \dots, Y_l that we call hidden variables, the complete set of variables is $\mathbf{X} = \{X\} \cup \mathbf{E} \cup \mathbf{Y}$, where X is the query variable. Then the inference query of X is $\mathbf{P}(X|\mathbf{e})$.

Any conditional probability can be computed by summing the factors from the full joint distribution. That is, $\mathbf{P}(X|\mathbf{e})$ can be computed by summing all the possible events of the hidden variables of the joint distribution.

$$\mathbf{P}(X|\mathbf{e}) = \alpha \mathbf{P}(X, \mathbf{E}) = \alpha \sum_{\mathbf{y}} \mathbf{P}(X, \mathbf{e}, \mathbf{y}) \quad (1)$$

Thus, we can use a Bayesian network to represent the full joint distribution. Furthermore, since:

$$P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i | \text{Parents}(X_i)) \quad (2)$$

Then we can write $P(x, \mathbf{e}, \mathbf{y})$ as products of the conditional probabilities of the Bayesian network. Because of that, we can compute a query via the sums of products of the conditional probabilities of the network. Report [3] shows how to represent a Bayesian network with sums and products.

After we have the resulting expression given by equations 1 and 2, we can use factor properties [4] to simplify the expression and avoid repeated calculations [5].

Since exact inference with Bayesian networks is, in the general case, intractable, it makes these methods in the real world unrealistic. Section 14.5 from [8] describes some methods for approximate inference. Darwiche (2009)[1] is also a great source for Bayes nets.

3 Bayesian learning

Referências

- [1] Adnan Darwiche. *Modeling and Reasoning with Bayesian Networks*. 1st Edition. Cambridge University Press, 2009.
- [2] Renato Lui Geh. *Aprendizado Automático de Sum-Product Networks (SPN)*. 2015. URL: <http://www.ime.usp.br/~renatolg/mac0215/doc/project/relatorio.pdf>.
- [3] Renato Lui Geh. *Modeling and Reasoning with Bayesian Networks: Compiling Bayesian Networks*. 1. 2015. URL: <http://www.ime.usp.br/~renatolg/mac0215/doc/reports/week1/relatorio.pdf>.
- [4] Renato Lui Geh. *Modeling and Reasoning with Bayesian Networks: Inference by Variable Elimination 6.1-6.5*. 2. 2015. URL: <http://www.ime.usp.br/~renatolg/mac0215/doc/reports/week2/relatorio.pdf>.
- [5] Renato Lui Geh. *Modeling and Reasoning with Bayesian Networks: Inference by Variable Elimination 6.6-6.9*. 3. 2015. URL: <http://www.ime.usp.br/~renatolg/mac0215/doc/reports/week5/relatorio.pdf>.
- [6] Renato Lui Geh. *Modeling and Reasoning with Bayesian Networks: Inference with Local Structure 13.1-13.3*. 4. 2015. URL: <http://www.ime.usp.br/~renatolg/mac0215/doc/reports/week8/relatorio.pdf>.
- [7] Renato Lui Geh. *Uma Introdução a Sum-Product Networks*. 5. 2015. URL: <http://www.ime.usp.br/~renatolg/mac0215/doc/reports/week9/relatorio.pdf>.
- [8] Stuart Russel e Peter Norvig. *Artificial Intelligence: A Modern Approach*. 3rd edition. Prentice Hall, 2010.