COS513: FOUNDATIONS OF PROBABILISTIC MODELS LECTURE 1

VIJAY KRISHNAMURTHY, YONGXIN TAYLOR XI

1. Motivation and Examples

COS513 is a graduate course focussed on the mathematical and algorithmic foundations of the field of probabilistic modeling. The availability of huge amounts of data has engendered many interesting questions from a wide variety of fields, such as:

(1) What are mechanisms underlying gene expression data? - Colon Cancer Research.
(2) How to predict prices of stocks and bonds from historical data? - Hedge fund dynamics.
(3) Given a list of movies that a particular user likes, what other movies would she like?- Netflix Prize.
(4) How to identify aspects of a patient's health that are indicative of disease? - Heart Disease Classification.
(5) Which documents from a collection are relevant to a search query? - Google Research.

Each of the above questions can be addressed with a carefully designed joint distribution of observed and hidden random variables. Application of probabilistic modeling to real world problems involves the following steps:

(1) Formulating questions about data.
(2) Design an appropriate joint distribution.
(3) Cast our questions on the computation on the joint.
(4) Develop efficient algorithms to perform or approximate the computations on the joint.

2. Joint Distributions

We now highlight some of the quantities that can be computed from a joint distribution, which can be furthered used to answer the questions posed in the previous section. Let (X_1, X_2, X_3, X_4) denote a vector of random variables and let $(x1, x2, x3, x4)$ denote a particular realization of these random variables. Relevant questions on the joint distribution include:

(1) What is the probability of a particular realization? For eg, $P(X_1 = x1, X_2 = x2, X_3 = x3, X_4 = x4)$.
(2) What are the marginal probability distributions that are implied? For eg, $P(X_1, X_2)$.
(3) What are some of the conditionals distributions that can be computed? For eg, $P(X_1, X_2|X_3)$.
(4) What are expected values of functions of random variables? For eg, $E[f(X_1, X_2)]$
(5) Questions related to independence of the random variables. For eg, is X_1, X_2 conditionally independent of X_3

2.1. **Complexity Issues.** Before we proceed to answer the above questions, it is important to keep in mind the huge storage and retrieval issues involved in the representation of the joint. Suppose we have 100 binary valued random variables, then a naive representation of the joint would translate into storing 2^{100} elements. Even assuming that it is feasible to store the joint distribution table, it is hopelessly time consuming and impractical to use it to answer questions such the marginal distribution of a particular variable.

2.2. **Course Objectives.** : In this class our objective would be to learn general purpose algorithms for computing the above functionals of the joint distribution, even as we gain some control over complexity. In this regard we will focus on learning to represent the joint distribution compactly using information related to independence and design approximate algorithms by cutting corners in the right places.

3. GRAPHICAL MODELS

We begin with the following definition of a graphical model.
$\underline{Def}$: Graphical models refer to a family of probability distributions defined in terms of a directed or undirected graph.

We now illustrate a few examples of important graphical models that we will encounter frequently ; We will cover the technical details in future lectures.

(1) Markov Chain

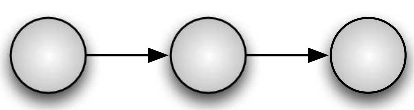

FIGURE 1. Markov Chain.

The shaded circles in Figure 1 (and in the following figures) represent observed variables.

(2) Hidden Markov Chain

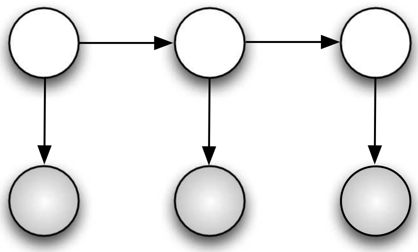

FIGURE 2. Hidden Markov Chain.

Figure.2 is also referred to as the Kalman Filter.

(3) Naive Bayes Classification

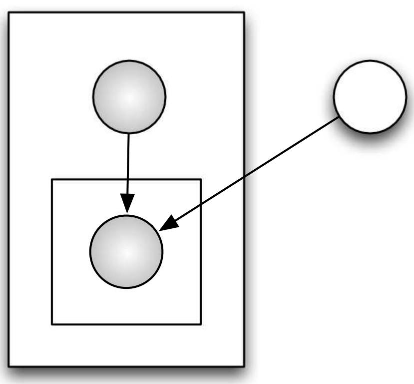

FIGURE 3. Naive Bayes Classifier.

(4) Mixture Model

Figure.4 also represents Factor Analysis.

(5) MRF, GRF, Ising Model

Figure.5 can also be used for Gaussian Random Field and Ising Model.

Graphical models (GM) have several advantages. For example, they:

(1) allow us to articulate structural assumptions about collections of random variables.
(2) provide general algorithms to compute conditionals, marginals, expectations and independencies, etc.

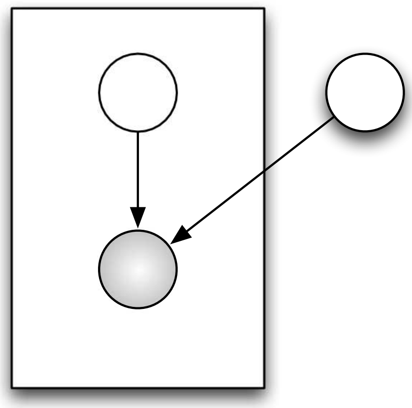

FIGURE 4. Mixture Model.

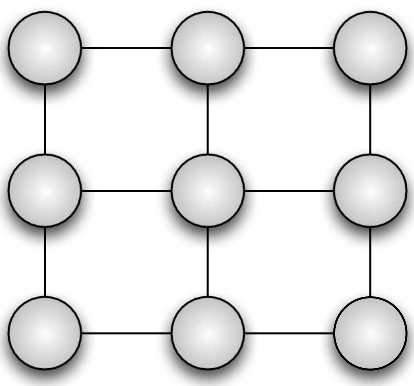

FIGURE 5. Markov Random Field.

(3) provide control over the complexity of these operations.
(4) decouple the factorization of the joint from its particular function form.

By using such modeling, we can connect models across disciplines, provide a natural language for expressing assumptions about the data, and provide a modular framework where visual embeddings from one model can be reused in other models.

4. CLASS LOGISTICS

Last but not the least, the grading of this class will consist of the following four components:

(1) Scribe notes, 15% (latex template available course website).
(2) Research report, 30% (3-5 pages due middle of the semester; sample topics could be model selection, dimensionality reduction, or regularized regression, etc.).
(3) Final project, 50% (8-10 pages; it could be theory-related or deriving strong meaningful results from data analysis).
(4) Class participation, 5%.