Contents

THESE NOTES form a concise introductory course on probabilistic graphical models[1]. They are based on Stanford CS228, and are written by Volodymyr Kuleshov and Stefano Ermon, with the help of many students and course staff. ⊕ You too may help make these notes better by submitting your improvements to us via Github.

This course starts by introducing probabilistic graphical models from the very basics and concludes by explaining from first principles the variational auto-encoder, an important probabilistic model that is also one of the most influential recent results in deep learning.

Preliminaries

1. Introduction: What is probabilistic graphical modeling? Overview of the course.

2. Review of probability theory: Probability distributions. Conditional probability. Random variables (*under construction*).

3. Examples of real-world applications: Image denoising. RNA structure prediction. Syntactic analysis of sentences. Optical character recognition (*under construction*).

Representation

1. Bayesian networks: Definitions. Representations via directed graphs. Independencies in directed models.

2. Markov random fields: Undirected vs directed models. Independencies in undirected models. Conditional random fields.

Inference

1. Variable elimination The inference problem. Variable elimination. Complexity of inference.

2. Belief propagation: The junction tree algorithm. Exact inference in arbitrary graphs. Loopy Belief Propagation.

3. MAP inference: Max-sum message passing. Graphcuts. Linear programming relaxations. Dual decomposition.

4. Sampling-based inference: Monte-Carlo sampling. Importance sampling. Markov Chain Monte-Carlo. Applications in inference.

5. Variational inference: Variational lower bounds. Mean Field. Marginal polytope and its relaxations.

Learning

1. Learning in directed models: Maximum likelihood estimation. Learning theory basics. Maximum likelihood estimators for Bayesian networks.

2. Learning in undirected models: Exponential families. Maximum likelihood estimation with gradient descent. Learning in CRFs

3. Learning in latent variable models: Latent variable models. Gaussian mixture models. Expectation maximization.

4. Bayesian learning: Bayesian paradigm. Conjugate priors. Examples (*under construction*).

5. Structure learning: Chow-Liu algorithm. Akaike information criterion. Bayesian information criterion. Bayesian structure learning (*under construction*).

Bringing it all together

1. The variational autoencoder: Deep generative models. The reparametrization trick. Learning latent visual representations.

2. List of further readings: Structured support vector machines. Bayesian non–parametrics.

Contents - Volodymyr Kuleshov