<u>10-708 – Probabilistic Graphical Models</u>

A pedagogical tool to track the key aspects of the lecture that the instructor emphasises, and as a checklist of things you have judged are important.

Some content is given little weight in the lectures for time-keeping purposes. They are listed in italics, and you must make a judgement on whether it is necessary to allocate time to understanding them.

Key areas to understand

Week 1

Lecture 1 – Introduction

- Understand that probabilistic graphical models as a language for representing complex probability distributions, reasoning under uncertainty and noise.
- Understand the intractability of working with full joint probability tables.
- Understand the benefits of probabilistic graphical models in this context.
- Understand the formal definition of a probabilistic graphical model.
- Understand that graph structure, topology, and conditional independence can be used to encode domain knowledge and yield representational cost-savings.

Lecture 2 – Directed Graphical Models

- Understand that Bayesian Networks/Directed Graphical Models have directed edges - causality relationships.
- Understand that Markov Random Fields/Undirected Graphical Models have undirected edges- correlation relationships.
- Understand the notational conventions of the course.
- Understand the HMM dishonest casino model in context of the knowledge engineering process.
- Understand the definition and properties of a Bayesian Network, the factorisation theorem, and the role of conditional independence.
- Understand the local structures and independencies.
- Understand the I-maps.
- Understand "explaining away"
- Understand the "d-separation criterion" on moralised ancestral graphs.
- Understand equivalence theorem.
- Understand "soundness" and "completeness" of d-separation with respect to a Bayesian Network factorisation law.
- Understanding of the concepts mathematically, but also be able to rehearse the semantics.

- *Understand the identifiability of local and global independence of Bayesian Networks via d-separation and the Bayes ball.*

Outstanding:

- *Jordan (2003): Understanding of Bayes ball algorithm.*
- *Jordan (2003): Understanding of explaining away.*
- *Complete this.*

Lecture 3 – Undirected Graphical Models

- Understand a motivation for undirected graphical models through P-maps.
- Understand that undirected graphical models have undirected edges – pairwise, non-causal, relationships.
- Understand the formal definition and properties of a Markov Random Field/Markov Network/UGM, in terms of the potential and partition function; and the Gibbs distribution.
- Understand the properties and interpretations of the clique potential and partition functions.
- Understand the formal definition of cliques, max-cliques, and sub-cliques and their relation to potential functions.
- Understand that the choice of clique size can yield representations of varying granularity in terms of the Gibbs distribution.
- Understand global Markov independencies – separation.
- Understand local Markov independencies – Markov blankets.
- Understand soundness, completeness of global Markov properties.
- Understand the theorems on the relation between local and global Markov properties.
- Understand the Hammersley-Clifford theorem.
- Understand the formal theorems on P-maps for UGMs.
- Be fluent in the semantics of the formalism to express relationships between distributions and graphs.
- Understand the unconstrained for of clique potential using exponentiated energy functions.
- Be familiar with UGM representations of Boltzmann Machines, Ising Models, Restricted Boltzmann Machines, and Conditional Random Fields.

Outstanding:

- *Minimal I-maps and P-maps*

Week 3

Lecture 4 – Exact inference and variable elimination

- Understand that queries on the graphical model can be viewed as statistical inference and learning/estimation problems.
- Understand evidence as an assignment of values to a set of random variables.
- Understand that we can query the likelihood, posterior, and maximum a posterior (MAP) or most probable assignment.
- Understand the associated examples, and how PGMs illustrate the explicit role of the context of an MAP query.
- Understand that computational complexity results of computing posteriors on arbitrary PGMs as NP-hard.
- Understand that inference approaches can be viewed as exact, or approximate.
- Understand the elimination algorithms for PGMs with a chain structure.
- Understand its application to examples such as the forward-backward algorithm for HMMs, undirected chains, and CRFs.
- Understand the role of the sum-product operation as providing a means of quantifying the complexity of the inference precisely.
- Understand the variable elimination algorithm for general PGMs at a high-level.
- Understand the specifics of variable elimination, such as the introduction of evidence, outcome of elimination, sum-product variable elimination.
- Understand how to evaluate the computational complexity of variable elimination and its notable determinants.

- Understand the variable elimination algorithm for more complex non-chain structured PGMs.
- Understand the graph-theoretic formulation of variable elimination.
- Understand that elimination can be viewed as message passing over an ordering over clique trees.
- Understand the nuances regarding complexity.

Outstanding:

<u>Lecture 5 – Parameter estimation in fully observed Bayesian Networks</u>

This was skipped in the Spring 2019 series; but is covered in Spring 2020 in the first half of Lecture 5. Only a proportion of the material in the slides is covered as the Spring 2020 concatenates the lecture content in Lecture 5 and 6 in Spring 2019 into Spring 2020 Lecture 5.

- Have knowledge of the main scenarios of learning/parameter estimation of PGMs, i.e. completely observed/partially observed/directed/undirected and estimation principles.
- Understand the distinction between PGM structure learning and parameter estimation.
- Understand the role of the generalised exponential family of distributions for parameter estimation in the context outlined.
- Understand that single and two-node graphical models form generalised building blocks for more complex graphical models.
- Understand the terms comprising the functional form of the exponential family distribution, such as the canonical parameter, sufficient statistic, and log normaliser.
- Understand the exponential family representation of the multivariate Gaussian, Multinomial.
- Understand the benefits of the exponential family representation.
- Understand the relation between the exponential family distributions and moments.
- Understand the relation between canonical and moment parameters, and the process of moment matching.
- Understand the definition and role of a sufficient statistic.
- Understand the relation between data, sufficient statistics and parameters under the frequentist and Bayesian paradigms; and the Neyman factorisation theorem.
- Understand how these various methods can give a unified approach to density estimation for single random variables/single node graphical models.
- Understand the definition of a Generalised Linear Model (GLM), and the terms comprising it, such as the response function, inverse transform function, and the role of the exponential family distribution.
- Understand the relation between maximum likelihood estimation (MLE) for canonical GLMs, online learning, batch learning.
- Understand the trade-offs between stochastic gradient ascent (SGA) and iteratively reweighted least squares (IRLS).
- Understand that MLE for general BNs, under certain conditions, can be decomposed analytically.
- Appreciate that the exponential family and GLMs, as one and two-node graphical models, can be used as generalised building blocks for more complex graphical models.

Outstanding:

<u>Week 4</u>

<u>Lecture 6 – Parameter estimation in partially observed GMs (POGMs)</u>

- Understand how parameter estimation is conducted for fully observed DGMs/BNs.
- Understand that in this setting, the log-likelihood function decouples into a set of local terms.
- Understand plate notation.
- Understand the process of ML parameter estimation for BNs with tabular CPDs.
- Understand ML estimation from HMMs in a fully observed, supervised learning setting.
- Understand ML in this setting estimates empirical probabilities/relative frequencies.
- Understand ML overfitting may be addressed via the inclusion of pseudo-counts; and amounts to the inclusion of a prior from a Bayesian perspective.
- Understand that parameter estimation in partially observed GMs involves iterating between inference tasks.
- Understand that inference can be viewed as a subroutine for parameter estimation.
- Understand the various senses in which variables can be unobserved, or latent.
- Understand the role latent variables play in mixture models, and Gaussian mixture models (GMMs).
- Understand that parameter estimation in POGMs is due to a coupling of likelihoods with latent variables.
- Understand the usefulness of the EM algorithm for ML estimation in the presence of unobserved latent variables.
- Understand a high-level description of the EM algorithm (inference of unobserved variables, given observed data and current parameter estimates, followed by updating parameter estimates).
- Understand the relation between EM for GMMs and soft-clustering K-means.
- Understand the isomorphy of MLE and EM via sufficient statistics.
- Understand the mathematical details of the components of the EM algorithm, such as the complete log-likelihood, incomplete log-likelihood, expected complete log-likelihood, and the posterior distribution over latent variables.
- Understand that the EM algorithm maximises the expected complete log-likelihood, a lower bound on the incomplete log-likelihood, via Jensen's inequality.
- Understand the EM algorithm in terms of free energy, entropy, and co-ordinate ascent.
- Understand applications of the EM algorithm for HMMs (Baum-Welch), BNs, and conditional mixture models (CMMs).
- Be aware of EM variants, and of its advantages and disadvantages.

<u>Lecture 7 – Parameter estimation for UGMs</u>

- Understand that in general, log-likelihoods for UGMs do not easily decompose into local terms, due to the presence of a normalisation constant over potential functions.
- Understand that inference/marginalisation will be required for parameter estimation of UGMs, even when all variables are fully observed.
- Reinforce understanding on the role of counts, empirical probabilities in MLE of PGMs.
- Reinforce understand that the Hammersley-Clifford theorem allows the specification of an UGM in terms of a Gibbs distribution and partition function.
- Understand how key sufficient statistics are computed – total counts and clique counts.
- Understand the issues the normalisation constant poses for parameter estimation.
- Understand the process of MLE for UGMs only produces a condition on clique marginals, and no means of attaining the ML parameters.
- Understand how the above yields a motivation for two workhorse algorithms: iterative proportional fitting (IPF); and generalised iterative scaling (GIS).

- Understand that IPF can be used for tabular potentials.
- Understand the derivation and properties of IPF as a fixed-point equation.
- Understand how IPF can be viewed as a co-ordinate ascent algorithm, and from the perspective of information theory as minimising Kullback-Leibler (KL) divergence.

- Appreciate that the spirit of the course is not only to be theoreticians, but to ultimately be concerned with the "messiness" of practical implementation, and its requisite considerations.
- Understand the infeasibility of using tabular potential functions in practice with a heuristic example – exponential representational cost.
- Understand that the granularities of tabular potential functions can be compressed using features encoding domain knowledge.
- Understand the relation between features, weights, micro-potentials and clique potentials, in context of the exponential family distribution and GLMs.
- Understand how linearisation and lower-bounds are invoked mathematically to set up a scaled-log-likelihood optimisation problem.
- Understand how GIS iteratively improves the lower-bound on scaled-log-likelihood, and its similarities with the IPF algorithm.

- Reinforce understanding of the functional form of the exponential family distribution.
- Understand the relation between the exponential family, maximum likelihood, and expectations of sufficient statistics.
- Understand that the exponential family distribution is the solution to a constrained, variational optimisation problem over either an entropy or KL-divergence objective – maximum entropy methods.
- Understand how the maximum-entropy principle (MEP) to parametrisation offers a dual perspective to MLE.
- Understand the label bias problem in HMMs, and the motivation for conditional random fields (CRFs).

- *Review CRF tutorial papers, inference and learning in CRFs*