جامعة محمد بن زايد للذكاء الاصطناعي
MOHAMED BIN ZAYED UNIVERSITY OF ARTIFICIAL INTELLIGENCE

Core Courses Syllabi

ML706 - Advanced Probabilistic and Statistical Inference

Title	Advanced Probabilistic and Statistical Inference
Code	ML706
Loading	4 Credit-hours
Prerequisites	ML 703 Probabilistic and Statistical Inference
Catalog Description	The study of probabilistic and statistical inference deals with the process of drawing useful conclusions about data populations or scientific truths from uncertain and noisy data. This course will cover some highly specialized topics related to statistical inference and their application to real-world problems. The main topics covered in this course are latent variable learning, kernel methods and approximate probabilistic inference strategies. This course will provide an in-depth treatment to various learning techniques (likelihood, Bayesian and max-margin) and numerous practical complexities (missing data, observed and unobserved confounding, biases) for performing inference.
Goal	During this course, the students will master some of the most important techniques for probabilistic and statistical inference and develop a broad understanding of the overall area. The specialized skill set developed in this course with be useful for making informed choices in analysing real-world data. The goal of this course is to master the state-of-the-art methods, promote discussions among students and motivate the students about the practical and scientific significance of reasoning about uncertainty. This course will provide the necessary background in frequentist and Bayesian approaches to statistical inference.
Contents	The course covers three core modules: **(I)** Learning with missing data (Expectation Maximization (EM) Algorithm for latent variable learning, Variational Bayesian EM algorithm, Variation Bayesian learning with Graphical Models), **(II)** Kernel Methods for Deep Learning (Kernel Machines, Deep Kernelized Networks, Structured Support Vector Machine), **(III)** Approximate Inference (Mean-field Approximate Inference, Loopy Belief Propagation, Generalized Belief Propagation, Expectation Propagation)
Recommended Textbooks	1. Bishop, C. Pattern Recognition and Machine Learning. Berlin: Springer-Verlag. ISBN: 0387310738 2. K. Murphy. Machine Learning: A Probabilistic Perspective. MIT, 2012.
Recommended References & Supplemental Material	1. Barber, D., 2012. Bayesian reasoning and machine learning. Cambridge University Press. 2. D. Koller and N. Friedman, 2009. Probabilistic Graphical Models: Principles and Techniques by, MIT Press. 3. D. J. C. MacKay, 2003. Information Theory, Inference and Learning Algorithms, Cambridge University Press. 4. S. Nowozin and C. Lampert, 2011. Structured Learning and Prediction in Computer Vision, Foundations and Trends in Computer Graphics and Vision Vol. 6, Nos. 3–4 (2010) 185–365

Teaching Week	Topics
1	**Lecture** • Introduction to "learning with latent variables" (particularly expectation maximization) • Summary of seminal papers on latent variable learning (Maximum likelihood, Variational and Classical Expectation maximization, Variational Bayes) • Following papers can be considered for group discussions and group study: - Dempster, A.P.; Laird, N.M.; Rubin, D.B. (1977). *"Maximum Likelihood from Incomplete Data via the EM Algorithm".* Journal of the Royal Statistical Society, Series B. 39 (1): 1–38. - Bernardo, J. M., et al. *"The variational Bayesian EM algorithm for incomplete data: with application to scoring graphical model structures."* Bayesian statistics 7 (2003): 453-464. - Attias, H. (2000). *"A variational bayesian framework for graphical models."* In Advances in neural information processing systems (pp. 209-215). • First week the group will discuss in detail the first seminal paper [Dempster et al., 1977] **Lab** • Project 1: Partial EM algorithm and the Viterbi training algorithm for its training, implementation, testing on an example machine learning problem as a use case.
2	**Lecture** • Reading group activity for the second seminal paper [Bernardo et al., 2003] on variational Bayesian Expectation-Maximization that addresses the learning with latent variables problem. • Discussion around [Dempster et al., 1977] and [Bernardo et al., 2003]. • Assessment of students understanding of [Dempster et al., 1977] with a graded quiz. **Lab** • Project 1: Partial EM algorithm and the Viterbi training implementation (continued).
3	**Lecture** • Reading group activity for the third seminal paper [Attias et al., 2000] on the variational Bayesian framework for graphical models that addresses the learning with latent variables problem. • Discussion around [Dempster et al., 1977], [Bernardo et al., 2003] and [Attias et al., 2000]. • Assessment of students understanding of [Bernardo et al., 2003] with a graded quiz. **Lab** • Project 1: Partial EM algorithm and the Viterbi training implementation (continued).

Teaching Week	Topics
4	**Lecture** • Feedback and peer review of the technical report prepared by the students around the review of the three above mentioned papers. • Assessment of students understanding of [Attias et al., 2000] with a graded quiz. • Discussion about one case-study in machine learning where learning with latent variables is required and how the variational EM can be applied. **Lab** • Project 1: Partial EM algorithm and the Viterbi training implementation (continued). • Preparation of presentation on the selected papers.
5	**Lecture** • Exam will be held instead of 1 scheduled lecture • Student presentations **Lab** • Complete project-1 work and presentation
6	**Lecture** • Introduction to “kernel methods in machine learning” (with a focus on support vector machines and deep learning) • Summary of seminal papers on kernel methods in machine learning (kernels; kernel methods for classification, regression and novelty detection; parameter learning for kernel machines, structured kernel methods) • The following papers can be considered for group discussions and group study: - Müller, K. R., Mika, S., Rätsch, G., Tsuda, K., & Schölkopf, B. (2001). *“An Introduction to Kernel-Based Learning Algorithms”.* IEEE Transactions on Neural Networks, 12(2), 181. (An overview on classical kernel methods) - Wilson, A. G., Hu, Z., Salakhutdinov, R., & Xing, E. P. (2016, May). *“Deep kernel learning.”* In Artificial Intelligence and Statistics (pp. 370-378). (the fusion of Deep learning and kernelization) - M. B. Blaschko and C. H. Lampert, *“Learning to localize objects with structured output regression,”* in European Conference on Computer Vision (ECCV), Springer, 2008 (the focus will be on Kernelized structured support vector machine) • During the first week the class group will discuss in detail the first seminal paper on an overview of kernel methods in machine learning [Muller et al., 2008]. **Lab** • Project 2: Kernel methods for deep learning, their training, implementation, testing on an example machine learning problem as a use case.

Teaching Week	Topics
7	**Lecture** • Reading group activity for the second seminal paper [Wilson et al., 2016] on deep kernel learning which addresses the combination of kernel learning with deep learning methods. • Discussion around [Muller et al., 2008] and [Wilson et al., 2016]. • Assessment of students understanding of [Muller et al., 2008] with a graded quiz. **Lab** • Project 2: Deep kernel learning implementation on a use-case (continued).
8	**Lecture** • Reading group activity for the third seminal paper [Blaschko et al., 2008] on the kernelized structured support vector machine, which addresses the problem domains that require structured reasoning. • Discussion around the [Muller et al., 2008], [Wilson et al., 2016] and [Blaschko et al., 2008]. • Assessment of students understanding of [Wilson et al., 2016] with a graded quiz. **Lab** • Project 2: Deep kernel learning implementation on a use-case (continued).
9	**Lecture** • Feedback and peer review of the technical report prepared by the students around the review of three above mentioned papers. • Assessment of students understanding of [Blaschko et al., 2008] with a graded quiz. • Discussion about one case-study in machine learning where learning with kernelized deep learning machines is required and how the discussed concepts can be applied. **Lab** • Project 2: Deep kernel learning implementation on a use-case (continued). • Preparation of presentation on the selected papers
10	**Lecture** • Exam will be held instead of 1 scheduled lecture • Student presentations **Lab** • Complete project-1 work and presentation

Teaching Week	Topics
11	**Lecture** • Introduction to “approximate inference in dynamical systems” (with a focus on probabilistic graphical models) • Summary of seminal papers approximate inference in probabilistic graphical models (deterministic approximate inference, variational approximate inference, discrete and continuous state markov models, belief propagation) • The following papers can be considered for group discussions and group study: - Yedidia, J. S., Freeman, W. T., & Weiss, Y. (2001). Generalized belief propagation. In Advances in neural information processing systems (pp. 689-695). - Xing, E. P., Jordan, M. I., & Russell, S. (2002, August). “*A generalized mean field algorithm for variational inference in exponential families*”. In Proceedings of the Nineteenth conference on Uncertainty in Artificial Intelligence (pp. 583-591). Morgan Kaufmann Publishers Inc. - Minka, T. P. (2001, August). “*Expectation propagation for approximate Bayesian inference*”. In Proceedings of the Seventeenth conference on Uncertainty in artificial intelligence (pp. 362-369). Morgan Kaufmann Publishers Inc. • First week the class group will discuss in detail the first seminal paper on generalized belief propagation [Yedidia et al., 2001]. **Lab** • Project 3: Variational Bounding Using KL divergence (mean-filed approximation) implementation and application to an example machine learning problem as a use case.
12	**Lecture** • Reading group activity for the second seminal paper [Xing et al., 2002] on a generalized mean field algorithm for variational inference in exponential families. • Discussion around [Yedidia et al., 2001] and [Xing et al., 2002]. • Assessment of students understanding of [Yedidia et al., 2001] with a graded quiz. **Lab** • Project 3: Mean-field approximate inference on a use-case (continued).
13	**Lecture** • Reading group activity for the third seminal paper [Minka et al., 2001] on the expectation propagation for approximate inference in probabilistic graphical models. • Discussion around the [Yedidia et al., 2001], [Xing et al., 2002] and [Minka et al., 2001]. • Assessment of students understanding of [Minka et al., 2001] with a graded quiz. **Lab** • Project 3: Mean-field approximate inference on a use-case (continued).

Teaching Week	Topics
14	**Lecture** • Feedback and peer review of the technical report prepared by the students around the review of three above mentioned papers. • Assessment of students understanding of [Minka et al., 2001] with a graded quiz exam. • Discussion about one case-study in machine learning where approximate inference in graphical models is required and how the discussed concepts can be applied in a practical problem. **Lab** • Project 3: Mean-field approximate inference on a use-case (continued). • Preparation of presentation on the selected papers
15	**Lecture** • Exam will be held instead of 1 scheduled lecture • Student presentations **Lab** • Complete project-1 work and presentation