The University of Connecticut **Fall 2018**
Dept. of ECE **KRP**

ECE 6141 : Neural Networks for Classification and Optimization

General Information

Instructor:

Dr. Krishna R. Pattipati
BoT Distinguished Professor and UTC Chair Professor in Systems Engineering
Room No.: ITEB 350
Phone/Fax: (860) 486-2890/5585
E-mail: **krishna.pattipati@uconn.edu**

Office Hours: Monday: 1:00-2:00PM; Tuesday: 11:00 AM – 12:00 Noon

Classes: Time: Monday, 6PM-9PM, Location: ROWE 318

Text: C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer, New York, 2006.
. (Optional) Sergios Theodoridis, Machine Learning, Academic Press, 2015.
(Optional) Kevin P. Murphy, Machine Learning: A Probabilistic Perspective, MIT Press, 2012.

Course Objective

This course is designed to provide students with a thorough understanding of the mathematical underpinnings of neural networks and graphical models, as well as the implementation and testing of various forms of neural networks and graphical models in software.

Course Outline

Lectures 1-2: **Introduction and Course Overview:**

- *Course Objectives*
- *What is Predictive Data Analytics?*
- *What is Machine Learning?*
- *Deep Learning and History of Neural Networks*
- *Getting to Know the Data: Data Visualization, Data Statistics, Data Scaling, Data Cleaning*
- *What kind of problems can be solved using Machine Learning?*
- *Basic Learning Tasks:* Density Estimation, Classification and Regression.
- *Generative versus Discriminative Learning*
- *Learning from Data:* MAP and ML Criteria, Likelihood-based Cost Functions, Canonical from of the Gradients of the Cost function, Optimization Algorithms
- *Graphical Models:* Naïve Bayes, Markov Chains, Hidden Markov Models, Factor Graphs, Markov Random Fields, Bayesian Networks

Lectures 3-5: **Statistical Inference and Probability Density Estimation**

- *Bayesian Hypothesis Testing:* Minimizing risk and probability of error, Decision rules for known distributions (Linear and Quadratic Discriminants, Logistic and Multinomial (Soft-max) Regression)
- *General Density Estimation:* Parametric methods, ML methods, Bayesian Inference, Sequential parameter estimation, Gaussian mixtures and the EM algorithm, Non-parametric methods
- *Performance Assessment of Decision Rules:* Confusion Matrix, Precision, Recall, Sensitivity, Specificity, Error Rates, Odds Ratio, Kappa, ROC)

- *Performance Assessment of Regression Models*: BIC, AIC, MDL and variants
- *Comparing Classification and Regression Models*

Lecture 6: **Decision-based Learning**

- *Classification as a Function Approximation Problem*
- *Approximating Posterior Probabilities*
- *Decision Based Learning*
- *Perceptron Convergence Theorem*
- *Optimal Learning Rate & Normalized Perceptron*
- *Extension to Multiple Classes*
- *Fuzzy Updates*
- *LVQ as a Perceptron Update*
- *Decision Trees (Method for constructing decision trees, Choosing Tests, Splitting Rules, Pruning Rules, Handling Missing Values, Extensions to Complex Tests, Boosted Decision Trees and Random forests)*

Lecture 7: **Regression-based Learning**

- *Linear and Nonlinear Regression Based Learning*: Least Squares (L2) Regression; Ridge Regression; L1 Regression; Combined L1 and L2 Regression
- *Nonlinear Optimization Techniques*
- *ADALINE (Adaptive Linear Element)*
- *Stochastic Convergence Analysis*
- *Support Vector Machines*
- *Kernel, Support Vector Machine (SVM) and Gaussian Process Regression*

Lectures 8-9: **Multi-layer Perceptrons and Deep Learning**

- *Multilayer Perceptrons:* Gradient Calculation via Back Propagation; Relationship to Calculus of Variations; Variants of the classic Back Propagation algorithm (Momentum term, Weight Decay, Quickprop, Learning Rate Adaptation, Incremental Gauss-Newton (Extended Kalman Filter))
- *MLP learning as an Optimization Problem:* Conjugate Gradient and Quasi-Newton Methods, Recursive Hessian Computation and Newton's Method, Levenberg-Marquardt Method.
- *MLP as Universal Function Approximator*
- *Practicalities:* type of nonlinearities, Initialization, batch versus recursive training, prior information
- *Managing Network Complexity:* Pruning Networks; Selecting the Number of Hidden Nodes; regularization; Incremental Network Construction
- *Deep Learning:* Need for Deep Architectures, Deep Convolutional Networks, Recurrent Neural Networks

Lecture 10: **Radial Basis Functions, Gaussian Processes and Relevance Vector Machines**

- *RBF Networks and Training*
- *Regularization Theory*
- *Relation to Kernel Regression*
- *Gaussian Processes*
- *Relevance Vector Machine*
- *Gaussian Unit-based Deep Learning*

Lecture 11: **Unsupervised Learning and Feature Extraction**

- *Projection Methods:* SVD, Projection Pursuit, probabilistic PCA
- *More Clustering Algorithms:* Dirichlet process mixture models, Spectral clustering, Hierarchical Clustering

- *Independent Component Analysis*
- *Self-organizing Map (SOM)*
- *Feature Selection*: Bayesian variable selection, sparse linear models, Greedy search, Discretization of Numeric Features: Entropy, Error-based and Unsupervised
- *Connections to Deep Learning:* Autoencoders and Stacked autoencoders

Lectures 12-13: Graphical Models for Machine Learning

- *Graphical Models:* Markov Chains, HMMs, Hybrid Systems, Factor Graphs, Markov Random Fields and Bayesian Belief Networks
- *The Sum-product Algorithm, Variational Inference, Loopy Belief Propagation, Monte Carlo Methods ((Importance Sampling, Markov Chain Monte Carlo (MCMC) Methods, Gibbs Sampling, Contrastive Divergence)*
- *Graphical Model Learning*
- Connection to Deep Learning: Deep Directed Networks, Deep Boltzmann Machines, Deep Belief Networks, Auto-Regressive Networks

Lecture 14: Neuro-dynamic Programming (unlikely to get to this!)

- *Principle of Optimality and Dynamic Programming (DP) Recursion*
- *Cost-to-go Approximations in DP*
- *Approximation Architectures*
- *Simulation and Training*
- *Neuro-dynamic Programming:* Q-learning, Temporal Difference Methods
- *Rollout policies*

Grading:

Homework/Project Assignments	40%
Mid-term Take Home Exam	25%
Review Paper Presentation	10%
Term Project	25%
Total	100%

Additional Information:

- Starting with **October 8** Lecture, each lecture will be divided into two parts. The first 2.5 hours will be used to present course materials, and the remaining 0.5-hour will be used for students to present reviews of recent journal publications.

- Paper reviews should be based on relevant and recent (2004 and up) journal articles from, e.g., Journal of Machine Learning Research, IEEE Trans. On Pattern Analysis and Machine Intelligence, Machine Learning (Journal), Neurocomputing, IEEE Trans. On Neural Networks, IEEE Trans. On Signal Processing, IEEE Trans. On Automatic Control IEEE Trans. On SMC, Pattern Recognition, Biological Cybernetics, Neural Computation, Statistical Science.

- Term projects can be performed in teams of two students on relevant neural networks topics. Topics could be related to research, or based on at least two recent journal articles. **Numerical implementation** and **testing** are a must.

- Term project **proposals** are due on **October 8**, presentations are scheduled for **Monday December 10** from 6 PM to 9 PM, and final reports are due on **Wednesday December 12 (final date because final grades are assigned on December 13th!)**.

- Programming can be done in any language.

References:

[ABU12] Y. S. Abu-Mostafa, M. Magdon-Ismail, H-T. Lin, Learning from Data: A Short Course, AMLbook.com, 2012.

[AMA98] S. Amari, N. Kasabov, *Brain-like computing and intelligent information systems*, Springer Verlag, Berlin, 1998.

[AUB96] J-P. Aubin, Neural Networks and Qualitative Physics, New York: Cambridge University Press, 1996.

[BIS95] C.M. Bishop, Neural Networks for Pattern Recognition, Oxford: Clarendon Press, 1995.

[BAT88] D.M. Bates and D.G. Watts, Nonlinear Regression Analysis and its Applications, New York: Wiley, 1988.

[BER99] M. Berthold, D. J. Hand, Intelligent Data Analysis, An Introduction, Springer-Verlag, 1999.

[BER96] Dimitri P. Bertsekas and John N. Tsitsiklis, Neuro-Dynamic Programming, Belmont, MA: Athena Scientific, 1996.

[BER97] D. Bertsimas and J. Tsitsiklis, Introduction to Linear Optimization, Athena, 1997.

[BIS97] C.M. Bishop, Neural Networks for Pattern Recognition, Clarendon Press, Oxford, 1997.

[BOY04] S. Boyd, and L. Vandenberghe, Convex Optimization , Cambridge, 2004.

[BRE84] L. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone, Classification and Regression Trees, Wadsworth, California, 1984.

[CAS01] G. Casella and R.L. Berger, Statistical Inference, Duxbury Press, 2001.

[CESA06] N. Cesa-Bianchi, and G. Lugosi, Prediction, Learning and Games, Cambridge University Press, 2006.

[CHE96] C.H. Chen (ed.), Fuzzy Logic and Neural Network Handbook, New York: McGraw-Hill, 1996.

[CHE98] V. Cherkassky and F. Mulier, Learning from data, concepts, theory and methods, John Wiley & Sons, New York, 1998.

[CHE94] V. Cherkasky, J.H. Friedman and H. Weschler (eds.), From Statistics to Neural Networks: Theory and Pattern Recognition Applications, Berlin:Springer, 1994.

[CHR00] N. Cristinanini and J. Shawe-Taylor, Support Vector Machines and other kernel-based learning methods, Cambridge University Press, Cambridge, UK, 2000.

[CIC93] A. Cichocki and R. Unbehauen, *Neural Networks for Optimization and Signal Processing*, John Wiley & Sons, New York, 1993.

[COV91] T.M. Cover and J.A. Thomas, Elements of Information Theory, New York:Wiley, 1991.

[DAS91] B. V. Dasarathy (ed.), Nearest Neighbor (NN) Norms: NN Pattern Classification Techniques, Los Alamitos, CA: IEEE Computer Society Press, 1991.

[DEV96] L. Devroye, L. Gyorfi, G.Lugosi, A Probabilistic Theory of Pattern Recognition, Springer-Verlag New York, Inc.1996.

[DEV82] P.A. Devijver and J. Kittler, *Pattern Recognition, a Statistical Approach*, Prentice Hall, Englewood Cliffs, London, 1982.

[DOMI09] P. Domingos and D. Lowd, Markov Logic: An Interface Layer for AI, Morgan & Claypool, 2009.

[DUD00] R.O. Duda, P.E. Hart and D. Stork, Pattern classification, John Wiley & Sons, NY, 2000.

[EFF10] B. Efron. Large-Scale Inference: Empirical Bayes Methods for Estimation, Testing, and Prediction, Cambridge, 2010.

[FER96] T. Ferguson, A Course in Large Sample Theory, Chapman & Hall/CRC.

[FRAS08] A. Fraser, Hidden Markov Models and Dynamic Systems, SIAM Press, 2008.

[FRE98] B.J. Frey, Graphical Models for Machine Learning and Digital Communication, Cambridge, MA: The MIT Press, 1998.

[FRI99] M. Friedman and A. Kandel, Introduction to Pattern Recognition, statistical, structural, neural and fuzzy logic approaches, World Scientific, Signapore, 1999.

[FUK90] K. Fukunaga, Introduction to Statistical Pattern Recognition, 2nd edition, Boston:Academic Press, 1990.

[GAL93] Stephen I. Gallant, Neural Network Learning and Expert systems, Massachusetts Inst. of Technology, Cambridge, Massachusetts, 1993.

[GEL03] A. Gelman, *et al*. Bayesian Data Analysis, Chapman & Hall/CRC, 2003.

[GOL96] G. Golub and C. Van Loan, Matrix Computations, Johns Hopkins, 1996.
[GOO16] I. Goodfellow, Y. Bengio and A. Courville, Deep Learning, MIT Press, 2016.
[GOL96] R.M. Golden, Mathematical Methods for Neural Network Analysis and Design, Cambridge, MA: The MIT Press, 1996.
[HAST09] T. Hastie, R. Tbishirani and J. Friedman, The Elements of Statistical Learning, Springer, 2009.
[HAN99] D. J. Hand, J. N. Kok and M. R. Berthold, Advances in Intelligent Data Analy*sis*, Springer Verlag, Berlin, 1999.
[HAY94] S. Haykin, Neural Networks: A Comprehensive Foundation, New York: MacMIllan, 1994.
[HAY91] S. Haykin, Adaptive Filter Theory, (3rd. Ed.), Englewood Cliffs, NJ: Prentice Hall, 1996.
[IRW96] J. David Irwin (ed.), The Industrial Electronics Handbook, Boca Raton, FL: CRC Press, 1996.
[JAYN03] T. Jaynes, Probability Theory: The Logic of Science, Cambridge University Press, 2003.
[JOR96] M.I. Jordan and C.M. Bishop, "Neural Networks", Chapter 23 in The Computer Science and Engineering Handbook, Allen B. Tucker (ed.), Boca Raton, FL: CRC Press, 1996.
[KOH95] T. Kohonen, Self-organizing Maps, Berlin:Springer, 1995.
[KOL09] D. Koller and N. Friedman, Probabilistic Graphical Models, Cambridge, MA: MIT Press, 2009.
[KOS92] B. Kosko, Neural networks for signal processing, Prentice-Hall, Englewood Cliffs, 1992.
[KUN93] S.Y. Kung, Digital Neural Networks, Prentice Hall, Englewood Cliffs, NJ, 1993.
[LEH95] E. Lehmann, Elements of Large-Sample Theory, Springer, 1995.
[LOO97] C.G. Looney, Pattern Recognition using Neural Networks, New York: Oxford, 1997.
[Mac03] David J.C. Mackay, Information, Theory, Inference and Learning Algorithms, Cambridge, UK: Cambridge University Press, 2003.
[MIC94] D. Michie, D.J. Spiegelhalter and C.C. Taylor (eds.), Machine Learning, Neural and Statistical Classification, New York: Ellis Horwood, 1994.
[MAA99] W. Maass and C. M. Bishop, Pulsed Neural Networks, MIT Press, Cambridge, 1999.
[MIN88] M. Minsky and S. Papert, Perceptron: An Introduction to Computational Geometry, (2nd edition), Cambridge, MA: The `MIT Press, 1988.
[MIT97] T.M. Mitchell, Machine learning, Mc Graw-Hill, New York, 1997.
[ORR97] G. B. Orr, K-R. Muller (editors), Neural Networks: Tricks of the Trade, Springer-Verlag Berlin Heildeberg, 1998.
[PAO89] Y.H. Pao, Adaptive Pattern Recognition and Neural Networks, Addison Wesley, Reading, Massachusetts, 1989.
[PEA88] J. Pearl, Probabilistic Inference in Intelligent Systems: Networks of Plausible Inference, San Mateo, CA: Morgan Kaufmann, 1988.
[PRAD10] R. Prado and M. West, Time Series: Modeling, Computation and Inference, CRC Press, 2010.
[RASM06] C. E. Rasmussen and J. Quinonero-Candela, Gaussian Processes for Machine Learning, MIT Press, 2006.
[RIP96] B.D. Ripley, Pattern Recognition and Neural Networks, Cambridge University Press, 1996.
[ROB05] C. Robert and G. Casella., Monte Carlo Statistical Methods, Springer, 2005.
[RUSS10] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, Prentice-Hall, 3rd edition, 2010.
[QUI93] J.R. Quinlan, C4.5: Programs for Machine Learning, San Mateo, CA: Morgan Kaufmann, 1993.
[RAS06] C. E. Rasmussen and C.K.I. Williams, Gaussian Processes for Machine Learning, MIT Press, Cambridge, MA, 2006.
[RUM86] D.E. Rumelhart and J.L. McClelland (eds.), Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume 1: Foundations, Cambridge, MA: the MIT Press, 1986.
[SCH99] B. Scholkopf, C. J. C. Burges, and A. J. Smola, Advances in Kernel Methods, *Support Vector Learning* MIT Press, Cambridge, 1999.
[SCH02] B. Scholkopf, and A. J. Smola, *Learning with Kernels*, MIT Press, Cambridge, 2002.

[SCH96] J. Schurmann, Pattern Classification: A Unified View of Statistical and Neural Approaches, New York: Wiley, 1996.

[SCH92] Robert Schalkhoff, Pattern Recognition, statistical, structural and neural approaches, John Wiley and Sons, New York, 1992.

[SCH90] G. Schafer and J. Pearl, Readings in Uncertainty Reasoning, San Mateo, CA: Morgan Kaufmann, 1990.

[SMO96] P. Smolensky, M. C. Mozer, and D. E. Rumelhart, *Mathematical Perspectives on Neural Networks*, Lawrence Erlbaum Associates, Inc. Mahwah, New Yersey, 1996.

[STOR04] D. G. Stork and E. Yom-Tov, *Computer Manual in MATLAB®to accompany Pattern Classification*, 2nd edition, John Wiley & Sons, New York, 2004.

[THE99] S. Theodoridis, K. Koutroumbas, Pattern recognition, Academic Press, 1999.

[THE15] S. Theodoridis, Machine Learning, Academic Press, 2015.

[VAP82] V. N. Vapnik, Estimation of Dependencies based on Empirical Data, New York: Springer, 1982.

[VAP98] V. N. Vapnik, Statistical Learning Theory, New York: Wiley, 1998.

[WAT85] S. Watanabe, Pattern Recognition, Human and Mechanical, John Wiley & Sons, New York, 1985.

[WEB99] A. Webb, Statistical pattern recognition, Oxford University Press Inc., New York, 1999.

[WEI91] S.M. Weiss and C.A. Kulikowski, Computer Systems that Learn, Morgan Kaufmann, San Mateo, California, 1991.

[WIN18] J. Winn, C. Bishop and T. Diethe, Model-based Machine Learning, http://www.mbmlbook.com/

[WHI92] H. White, Artificial Neural Networks: Approximation and Learning Theory, Oxford: Blackwell, 1992.