ECE 8230 Pattern Recognition: Statistical and Geometrical Approach, Winter 2013

Location: M, F: E2-399, 6:PM - ?
Instructor: Dr. Mirek Pawlak, Room E1-528, tel. 474-8881
Email: Miroslaw.Pawlak@ad.umanitoba.ca
Textbooks

1. Pattern Recognition, S. Theodoridis and K. Koutroumbas, Academic Press, 2009.
2. Pattern Classification, R.Duda, P.E.Hart and D.G.Stork, Wiley, 2001
   http://www.rii.ricoh.com/~stork/DHS.html

3. Lecture Notes
Course Mark

- Assignments - 20%
- Seminar/Term paper - 40%
- Final Exam - 40%

Assignments will be corrected, graded, and returned as soon as possible. Late homework will not be accepted and will receive a grade 0. The process of actively struggling with an assignment is one of the most important educational experiences you will have in this subject. Your main incentive for consistently doing the assignments should be to enhance your comprehension of the material. The assignments are graded to point out misconceptions and to give you an objective evaluation of your understanding. You can discuss the problems with other students, but do not submit someone else's solution.

Moreover students will be asked to read some material (papers, books, information coming from internet) for selected problems. They will require to write a 1-2 pages review of the provided topic.
• The purpose of the project (term paper) is to make you familiar with at least one application of pattern recognition reported in the literature. Alternatively one can select more theoretical topic which is not covered in the course. The project will involve choosing a particular problem (theory or application), reading a few papers on this topic in the current literature, and writing a 10 -15 page report.
If the topic is in the area of application of PR then report should emphasize the following design issues: the kind of pattern recognition problem (supervised, unsupervised, partially supervised), number of pattern classes, data structure (number of training and test samples, number and type of features used), decision rules applied, learning procedure used, validation techniques used, and the recognition accuracy and complexity.

Note: Topics for projects MUST come from internationally recognized journals which list is given below.

1. IEEE Transactions on Pattern Analysis and Machine Intelligence
2. Pattern Recognition Journal
3. Neural Computation
4. IEEE Trans. on Neural Networks
5. Pattern Recognition Letters
6. Pattern Analysis and Applications
7. IEEE Transactions on Information Theory
8. IEEE Transactions on Systems, Man and Cybernetics
9. IEEE Transactions on Acoustics, Speech and Signal Processing
9. Machine Learning
10. Journal of Machine Learning Research
11. Journal of Complexity
12. IEEE Transactions on Geoscience and Remote Sensing
13. Neural Networks
14. International Journal of Pattern Recognition and Artificial Intelligence
15. Journal of Classification
16. IEEE Trans. on Image Processing
17. JASA, Biometrika, Annals of Statistics, Journal of Multivariate Analysis - high quality statistical journals
Prerequisites: This course assumes that students have a working knowledge of probability theory, linear algebra, optimization methods, basic estimation techniques and other statistical topics on the level of introductory courses in statistics. Common sense, however, is the most required prerequisite.

Course Objective: This course will give a detailed overview of statistical, geometrical and approximation based classification techniques. Known as machine learning/data mining in computer science and artificial intelligence, pattern recognition in engineering, and discriminant analysis in statistics, this is a diverse and interesting area, with applications in science, industry and finance. The course will cover methods from linear classifiers through various classes of more flexible models to fully nonparametric techniques. Feature generation, selection and extraction techniques will be examined. Furthermore recent advances on combination of weak classifiers will be also overviewed. Both supervised and unsupervised learning methods will be discussed. The course attempts to focus on fundamental issues rather than application-dependent details, although the latter will not be completely ignored.
Pattern Recognition

Machine Learning

Data Mining

Discriminant Analysis
The Top Ten Algorithms in PR, ML, DM

1. CART/C4.5
2. K-Means
3. SVM
4. EM
5. Apriori
6. PageRank
7. AdaBoost
8. k-NN
9. Naive Bayes
10. Deep Learning Algorithm
Course Outline

1. **Overview of Learning and Pattern Recognition**

2. **Discriminant Functions - Linear Methods**
   2.1 **Fisher Method**
   2.2 **Linear Discriminant Functions - Perceptron Learning Rules**
   2.3 **Support Vector Classifiers**
   2.4 **Sparse Kernel Machines**
   2.5 **Nonlinear Discriminant Functions (Neural Nets-Deep Learning)**
   2.6 **On-Line Learning**

3. **Statistical Decision Theory**
   3.1 **Bayes Rule and Bayes Risk**
   3.2 **Bayes Rules for Specific Models: Normal Models, Binary Data, Logistic Discrimination, Naive Bayes**
   3.3 **Bayesian Networks**
4. Parametric Modeling

4.1 Normal Models
4.2 Logistic Regression

5. Nonparametric Modeling

5.1 Kernel Classifiers
5.2 Nearest Neighbor Decision Rules
5.3 Tree Classifiers

6. Model Assessment and Fusion

6.1 Asymptotic Accuracy
6.2 Error Estimation
6.3 Classifier Fusion: Bagging, Boosting, Arcing
6.4 Random Forest
7. Feature Selection and Extraction

7.1 Feature Generation
7.2 Data Visualization
7.3 Reduction of Dimensionality: Principal Components; Nonlinear Methods
7.4 Feature Selection

8. Unsupervised Learning

8.1 Mixture Models and EM Algorithms
8.2 Clustering: k-Means Algorithms
8.3 Clustering: Spectral Methods
9*. Context Classifiers

9.1 Speech Recognition and Hidden Markov Models
9.2 Spatial Context and Image Analysis
9.3 Classification of Time Series Signals

10*. Miscellaneous Topics

• methods for scaling up supervised learning algorithms
• reinforcement learning
• learning stochastic models
• incomplete data
• probabilistic teacher
• invariant classifiers
• information theory and pattern recognition; minimum length principle
LITERATURE ON PATTERN RECOGNITION

A Books on Statistical and Geometrical Approaches to PR

1. Sebestyen, G. (1962). Decision-Making Processes in Pattern Recognition. Macmillan, NY.
2. Nilsson, N.J. (1965). Learning Machines: Foundations of Trainable Pattern Classifying Systems. McGraw-Hill (reprinted in 1990 by Morgan Kaufmann).
3. Minsky, M. and Papert, S.A. (1969 and 1988). Perceptron: An Introduction to Computational Geometry. The MIT Press. (1988- expanded edition of the 1969 book).
4. Duda, R.O. and Hart, P.E. (1973). Pattern Classification and Scene Analysis. Wiley
5. Fukunaga, K. (1972 and 1990). Introduction to Statistical Pattern Recognition. Academic Press, (there is a 1990 revised edition).
6. Patrick, E. (1972). Fundamentals of Pattern Recognition. Prentice-Hall.
7. Meisel, W.S. (1972). Computer Oriented Approaches to Pattern Recognition. Academic Press.
8. Chen, C.H. (1973). Statistical Pattern Recognition. Hayden.
9. Tou, J. and Gonzalez, R. (1974). Pattern Recognition Principles. Addison Wesley.
10. Young, T.Y. and Calvert, T.W. (1974). Classification, Estimation and Pattern Recognition. Elsevier.
11. Lachenbruch, P.A. (1975). Discriminant Analysis. Hafner Press.
12. Chien, Y.T. (1978). Interactive Pattern Recognition. Dekker.
13. Batchelor, B. (1978). Pattern Recognition: Ideas in Practice. Plenum Press.
14. Agarwala, A.K., ed. (1977). Machine Recognition of Patterns. IEEE Press.
15. Fu, K.S., ed. (1982). Applications of Pattern Recognition. CRC Press.
16. Sklansky, J. and Wassel, G.N. (1981). Pattern Classifiers and Trainable Machines. Springer-Verlag.
17. Bow, S.T. (1984). Pattern Recognition. Dekker.
18. Devijver, P. and Kittler, J. (1982). Statistical Pattern Recognition. Prentice-Hall.
19. Krishnaiah, P. and Kanal, L., eds. (1982). Handbook on Statistics, Classification, Pattern Recognition and Reduction of Dimensionality. Vol. 2, North Holland.
20. Fu, K.S. and T.Y. Young, T.Y., eds. (1985). Handbook of Pattern Recognition and Image Processing. Academic Press.
21. Watanabe, S. (1985). Pattern Recognition: Human and Mechanical. Wiley.
22. James, M. (1985). Classification Algorithms. Wiley.
23. Fu, K.S. and Yu, T.S. (1982). Statistical Pattern Classification Using Contextual Information. Wiley.
24. Hand, D. (1981). Discrimination and Classification. Wiley.
25. Hand, D. (1982). Kernel Discrimination Analysis. Wiley.
26. Vapnik, V. (1982). Estimation of Dependences Based on Empirical Data. Springer.
27. Oja, E. (1983). Subspace Methods of Pattern Recognition. Wiley.
28. Breiman, L. and Friedman, J.H., Olshen, R., and Stone, C. (1984). Classification and Regression Trees. Woolsworth.
29. Fu, K.S., ed. (1984). VLSI for Pattern Recognition and Image Processing. Springer Verlag.
30. Devijver, P. and Kittler, J., eds. (1987). Pattern Recognition Theory and Applications. Springer Verlag.
31. Choi, S.C., ed. (1986). Statistical Methods of Discrimination and Classification-Advances in Theory and Applications. Pergamon Press.
32. Therrien, C. (1989). Decision Estimation and Classification: An Introduction to Pattern Recognition and Related Topics. Wiley.
33. Pao, Y.H. (1989). Adaptive Pattern Recognition and Neural Networks. Addison-Wesley.
34. Aivazyan, S.A., Buchstaber, V. M., Yenyukov, I.S., and Meshalkin, L.D. (1989). Classification and Reduction of Dimensionality. Finansy and Statistika (in Russian).
35. Dasarathy, B.V. (1991). NN Pattern Classification Techniques. IEEE Computer Society Press.
36. Weiss, S. and Kulikowski, C. (1991). Computer Systems that Learn. Morgan Kaufmann.
37. Schalkoff, R.J. (1992). Pattern Recognition: Statistical, Structural and Neural Approaches. Wiley.
38. McLachlan, G.J. (1992). Discriminant Analysis and Statistical Pattern Recognition. Wiley.
39. Rabiner, L. and Juang, B.H. (1993). Fundamentals of Speech Recognition. Prentice Hall.
40. Kearns, M.J. and Vazirani, U.V. (1994). An Introduction to Computational Learning Theory. The MIT Press.
41. Anzani, Y. (1992). Pattern Recognition and Machine Learning. Academic Press.
42. Nadler, M. and Smith, E.P. (1993). Pattern Recognition Engineering. Wiley.
43. Bishop, C.M. (1995). Neural Networks for Pattern Recognition. Oxford Press.
44. Vapnik, V. (1995). The Nature of Statistical Learning Theory. Springer.
45. M.H. Hassoun (1995). Fundamentals of Artificial Neural Networks. The MIT Press.
46. Hagan, M., Demuth, H. and Beale, M. (1995). Neural Network Design. PWS Publishing.
47. Ripley, B. (1996). Pattern Recognition and Neural Networks. Cambridge University Press, 1996.
48. Devroye, L., Gyorfi, L., and Lugosi, G. (1996). A Probabilistic Theory of Pattern Recognition. Springer.
49. Gose, E., Johnsonbaugh, R. and Jost, S. (1996). Pattern Recognition and Image Analysis. Prentice Hall.
50. Ballard, D.H. (1997). An Introduction to Natural Computation. The MIT Press.
51. Looney, C.G. (1997). Pattern Recognition Using Neural Networks. Oxford Press.
52. Michell, T.M. (1997). Machine Learning. McGraw-Hill.
53. Hand, D. (1997). Construction and Assessment of Classification Rules. Wiley.
54. Vidyasagar, M. (1997). A Theory of Learning and Generalization. Springer-Verlag.
55. Jelinek, F. (1997). Statistical Methods for Speech Recognition. The MIT Press.
56. Vapnik, V. (1998). Statistical Learning Theory. Wiley.
57. Theodoridis, S. and Koutroumbas, K. (1998). Pattern Recognition. Academic Press.
58. Cherkasky, V. and Mulier, F. (1998). Learning from Data. Wiley. 2nd Edition, 2007.
59. Sutton, R.S. and Barto, A. (1998). Reinforcement Learning. The MIT Press.
60. Frey, B.J. (1998). Graphical Models for Machine Learning and Digital Communication. The MIT Press.
61. Anthony, M. and Bartlett, P. (1999). Neural Network Learning: Theoretical Foundations. Cambridge Press.
62. Reed, R. and Marks, R.J. (1999). Neural Smithing. The MIT Press.
63. Webb, A. (1999). Statistical Pattern Recognition. Oxford University Press., 2-nd edition, 2002.
64. Principe J., Euliano, N. and Lefebvre, W.C. (2000). Neural and Adaptive Systems. Wiley.
65. Cristianini, N. and Shawe-Taylor, J. (2000). An Introduction to Support Vector Machines. Cambridge Press.
66. Smola, A., Bartlett, P., Scholkopf, B. and Schuurmans, D., eds. (2000). Advances in Large Margin Classifiers. The MIT Press.
67. Duda, R., Hart, P.E. and Stork, D.G. (2001). Pattern Classification. Wiley.
68. Raudys, S.(2001). Statistical and Neural Classifiers. Springer.
69. Hastie.T, Tibshirani,R. and Friedman, J. (2001, 2009). The Elements of Statistical Learning. Springer.
70. Herbrich, R.. (2002). Learning Kernel Classifiers. The MIT Press.
71. Scholkopf, B and Smola, A. (2002). Learning with Kernels. The MIT Press.
72. Alpaydin, E. (2004, 2010). Introduction to Machine Learning. The MIT Press.
73. Shawe-Taylor, J. and Cristianini, N. (2004). Kernel Methods for Pattern Analysis, Cambridge.
74. Kuncheva, L.I. (2004). Combining Pattern Classifiers. Wiley.
75. Neapolitan, R. (2004). Learning Bayesian Networks. Prentice Hall.
76. van der Heijden, Duin, R.P.W. and Tax, D.M.J. (2004). Classification, Parameter Estimation and State Estimation. Wiley.
77. Chapelle, Scholkopf, B., and Zien, A. (2006). Semi-Supervised Learning. The MIT Press.
78. Bishop, C.M. (2006). Pattern Recognition and Machine Learning. Springer.
79. Cuker, F. and Zhou D.X. (2007). Learning Theory. Cambridge.
80. Larose, D. (2005, 2006). An Introduction to Data Mining, DM Methods and Models. Wiley.
81. Bianchi, N.C. and Lugosi, G. (2006). Prediction, Learning, and Games. Cambridge.
82. Steinwart, I. and Christmann, A.(2008). Support Vector Machines, Springer.
83. Clarke, B., Fokoue, E., and Zhang, H.H. (2009). Principles and Theory for Data Mining and Machine Learning. Springer.
84. Berk, R.A. (2008). Statistical Learning from a Regression Perspective. Springer.
85. Izenman, A. J. (2008). Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning. Springer.
86. Marsland, S. (2009). Machine Learning: An Algorithmic Perspective. Chapman and Hall.
87. Krzanowski, W. and Hand, D. (2009) ROC Curves for Continuous Data, CRC Press.
88. Lee, J.A. and Verleysen, M. (2010). Nonlinear Dimensionality Reduction. Springer.
89. Kulkarni, S. and Harman. G. (2011). An Elementary Introduction to Statistical Learning Theory, Wiley.
90. Webb, A.R. and Copsey, K.D. (2011). Statistical Pattern Recognition. Wiley.
91. Rogers, S. and Girolami, M. (2011). A First Course in Machine Learning. Chapman and Hall.
92. Murphy, K. (2012). Machine Learning: A Probabilistic Perspective, The MIT Press.

• Good review paper: Kulkarni, S., Lugosi, G. and Venkatesh, S. (1998). Learning Pattern Classification. IEEE Information Theory, vol. 44, pp.2178-2206.

• Conference Proceedings: Advances in Neural Information Processing Systems (NIPS). The MIT Press.
B Syntactic Methods

1. Fu, K.S. (1974). Syntactic Methods in Pattern Recognition. Academic Press.
2. Pavlidis, T. (1977). Structural Pattern Recognition. Springer Verlag.
3. Gonzalez, R. and Thomason, R. (1978). Syntactic Pattern Recognition. Addison Wesley.
4. Fu, K.S.(1982). Syntactic Pattern Recognition and Application. Prentice-Hall.
5. Pavel, M.(1989). Fundamentals of Pattern Recognition. Dekker

C Unsupervised Learning

1. Hartigan, J. (1975). Clustering Algorithms. Wiley.
2. Everitt, B. (1978). Graphical Techniques for Multivariate Data. North Holland.
3. Everitt, B.(1977, 2010). Cluster Analysis.Heinemann Educational Books.
4. Anderberg, M.R.(1973). Cluster Analysis for Application. Academic Press.
5. Jain, A.K. and Dubes, R.(1980). Algorithms for Clustering Data. Prentice-Hall, 1988.
6. Titterington, D.M., Smith, A.F.M., and Makov, U.E.(1982). Statistical Analysis of Finite Mixture Distributions. Wiley.
7. Kauffman, L. and Rousseeuv, P.(1990). Finding Groups in Data. Wiley.
8. Kohonen, T. (1988). Self-Organization and Associative Memory. Springer.
9. McLachlan, G.J. (1997). Recent Advances in Finite Mixture Models. Wiley.

D Applications of PR

1. Jurs, P.C. and Isenhour, T.L. (1975). Chemical Applications of Pattern Recognition. Wiley.
2. Bayer, U. (1985). Pattern Recognition Problems in Geology and Paleontology. Lectures Notes in Earth Science, vol.2, Springer.
3. Rabiner, L. and Juang, B.H. (1993). Fundamentals of Speech Recognition. Prentice Hall.

**E Statistical Books Useful for PR**

1. Silverman, B.W. (1986). Density Estimation. Chapman and Hall.
2. Scott, D.W. (1992). Multivariate Density Estimation. Wiley.
3. Devroye, L. and Gyorfi, L. (1985). Nonparametric Density Estimation. Wiley.
4. Efromovich, S. (1999). Nonparametric Curve Estimation. Springer.
5. Loader, C. (1999). Local Regression and Likelihood. Springer.
6. Efron, B. and Tibshirani, R.J. (1993). An Introduction to the Bootstrap. Chapman and Hall.
7. McLachlan, G. and Krishnan, T. (1997). The EM Algorithm and Extensions. Wiley.
8. Mardia, K.V., Kent, J.T., and Bibby, J.M. (1979). Multivariate Analysis. Academic Press.
9. Johnson, R. and Wichren, D. (1988). Applied Multivariate Statistical Analysis. Prentice Hall.
10. Flurry, B. (1997). A First Course in Multivariate Statistics. Springer.
11. Buhlmann, P. and van der Geer, S. (2011). **Statistics for High-Dimensional Data**, Springer.

**F Journals**

1. IEEE Transactions on Pattern Analysis and Machine Intelligence
2. Pattern Recognition Journal
3. Neural Computation
4. IEEE Trans. on Neural Networks
5. Pattern Recognition Letters
6. Pattern Analysis and Applications
7. IEEE Transactions on Information Theory
8. IEEE Transactions on Systems, Man and Cybernetics
9. IEEE Transactions on Acoustics, Speech and Signal Processing
10. Machine Learning ←
11. Journal of Machine Learning Research ←
12. Journal of Complexity
13. Neural Networks
14. International Journal of Pattern Recognition and Artificial Intelligence
15. Journal of Classification
16. IEEE Trans. on Image Processing
17. JASA, Biometrika, Annals of Statistics, Journal of Multivariate Analysis - high quality statistical journals

Data for PR

http://www.ics.uci.edu/~mlearn/MLRepository.html