2012 Haptics Symposium

Tutorial on Machine Learning Methods for Human-Computer Interaction

Extended Outline

Kerem Altun

Postdoctoral Fellow
Department of Computer Science
University of British Columbia
Vancouver, B.C., Canada

kaltun@cs.ubc.ca

The following is an extended outline for the machine learning tutorial session organized as part of 2012 Haptics Symposium. I am not able to provide the tutorial slides at the time of this writing, but I am happy to send you the slides after the tutorial if you contact me at my e-mail address above.

Machine learning is a very broad field, and arguably, even people who spend a lifetime doing research in the field cannot learn all the aspects of it. In a 3-hour tutorial session, it is only possible to provide an introduction to the subject. Since the audience is assumed to have little or no background knowledge, I narrowed down the broad area of machine learning. Below I list various taxonomies of machine learning and state the tutorial scope with respect to each taxonomy. Machine learning algorithms can be categorized as:

1. ***Classification*** or ***regression*** algorithms, depending on the task and the output. For classification the output of the algorithm is discrete, for regression it is continous. We consider classification tasks only.
2. ***Supervised*** or ***unsupervised*** learning algorithms, depending on whether the observed samples are labeled beforehand or not. We consider supervised learning only.
3. ***Generative*** or ***discriminative*** approaches, depending on whether a probabilistic model is built from the observed samples or not. We consider the basic methods for both approaches.

In other words, we consider only classification (or ***pattern recognition***) methods with labeled data. Pattern recognition can also be categorized in four approaches [1]: ***template matching***, ***statistical classification***, ***syntactic*** or ***structural matching***, and ***neural networks***. In this tutorial we only consider statistical classification methods.

Regarding the approach to the subject and the notation, I mostly follow the book by Duda et al. [2] which is one of the classical textbooks in the field. I also used a few other books [3,4].

The concepts introduced in the tutorial are supported by examples either from artificial or real data. As examples, I use the excellent introductory scenario from [2], artificial bivariate Gaussian dataset from [1], real inertial sensor data from my previous work on human activity recognition [5], and real touch sensing data from the recent work of our group [6].

The tutorial is roughly organized in sections which I summarize briefly below:

1. **Introduction and Motivation.** I introduce the concept of machine learning and pattern recognition and set the scope of the tutorial. Then I discuss the example problems on pattern recognition and how to solve these problems with the participants.

2. **Probability theory review.** I recite the axioms of probability and the definition of conditional probability. Through an example I introduce the basic pattern recognition problem. I briefly review discrete and continuous random variables, and the univariate and multivariate Gaussian distributions.

3. **The Generative Approach to Pattern Recognition.** Through examples, I define prior and posterior probabilities as used in a classification problem, and introduce Bayesian decision theory when class-conditional densities are known. For the case when they are not known I give examples of parametric class-conditional probability density function estimation for the multivariate Gaussian case. Non-parametric models are also considered: histogram and kernel density estimation methods.

4. **The Discriminative Approach to Pattern Recognition.** I explain how the introduced generative methods result in decision boundaries and decision regions. Then I present different approaches to the linear decision boundary problem. Other decision boundaries are also considered with two popular methods: the nearest neighbor and *k*-nearest neighbor classifiers.

5. **Feature Reduction and Selection.** The examples for the previous sections make use of few features and the performance can not be expected to be good. But using too many features causes degradation in the performance as well – this is called the "curse of dimensionality." I present the simplest dimensionality reduction method – principal components analysis (PCA). Through examples, I explain how the classification performance increases with selecting many features and then reducing them with PCA. Sequential feature selection methods can also be mentioned if time permits.

6. **Statistical Cross Validation.** Usually the labeled samples used for training a classifier should not be used to test its performance, since the generalization ability of the classifier can not be evaluated correctly this way. However, obtaining different training and test data could be inconvenient. Here I present the usual cross-validation methods – repeated random subsampling, *K*-fold cross validation, leave-one-out cross validation.
7. **Evaluating Classifier Performance.** If time permits, I will provide a more detailed analysis of the two-class confusion matrix. I will define type I and type II errors, and how they change according to a decision threshold. Based on this information, I will introduce receiver operating characteristics curves and their interpretation.
8. **Further Resources.** If time permits, I will briefly discuss the extensions of the introduced methods – Gaussian mixture models, quadratic and other nonlinear discriminants, support vector machines. I will also provide some guidelines on combining different classifiers.

References

[1] A. K. Jain, R. P. W. Duin, J. Mao, "Statistical pattern recognition: a review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1):4—37, January 2000.

[2] R. O. Duda, P. E. Hart, D. G. Stork, *Pattern Classifiction*, 2nd ed., Wiley-Interscience, 2001.

[3] A. R. Webb, *Statistical Pattern Recognition*, 2nd ed., John Wiley & Sons, West Sussex, England, 2002.

[4] V. N. Vapnik, *The Nature of Statistical Learning Theory*, 2nd ed., Springer-Verlag New York, Inc., 2000.

[5] K. Altun, B. Barshan, O. Tunçel, "Comparative study on classifying human activities with miniature inertial and magnetic sensors," *Pattern Recognition*, 43(10):3605—3620, October 2010.

[6] A. Flagg, D. Tam, K. MacLean, R. Flagg, "Conductive fur sensing for a gesture-aware furry robot," *Proceedings of IEEE 2012 Haptics Symposium*, March 4-7, 2012, Vancouver, B.C., Canada.