Course Overview

Course Syllabus

Please see syllabus handout. The syllabus for EE565 Pattern Recognition and Machine Learning is also available at:

http://wordpress.nmsu.edu/pdeleon/teaching/ee565/syllabus

Course Outline

Please see outline handout. The outline for EE565 Pattern Recognition and Machine Learning is also available at:

http://wordpress.nmsu.edu/pdeleon/teaching/ee565/outline

Lecture Outline

Reading: Chapter 1

- Introduction

Reading: Appendix A

- MNIST and Old Faithful data sets

1 Introduction

Some useful and interesting applications of PRML:

- Classification of email as spam or non-spam (ham)
- Automatic identification of a unknown person based on fingerprints, voice, DNA, iris (biometrics)
- Recognition of handwriting (OCR) and hand-written zip codes (USPS)
- Identification of an unknown song (Shazam, SoundHound)
- Recommendation of music (Pandora, iTunes Genius) and movies (NetFlix)
- Identification of people you may know (Facebook)
- Automatic speech recognition (speech-to-text) (Nuance's Dragon dictation)
- Identification of spoken language (DARPA)
- Clustering of similar faces in a collection pictures (iPhoto)
- Recognition and translation of road signs (Google)

Definition: The field of *pattern recognition* is concerned with the *automatic discovery of regularities* in data through the use of computer algorithms and with the use of these regularities to take actions such as classifying the data in different categories.

Consider the example of recognizing handwritten digits (Fig. 1). The goal is to build a machine (computer program) that will take an input vector $\mathbf{x}$, whose elements are the stacked 28×28 pixel values, i.e. a 784×1 vector and produce the identity of the digit as the output. This is a nontrivial problem due to the wide variability of people's handwriting.

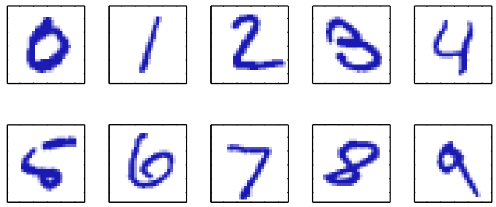

Figure 1:

Training, Testing, and Generalization

One solution to this problem uses machine learning in which a large set of N handwritten digits called a *training set* is used to tune the parameters of a model. The digit in each of the training images is known in advance by inspection, i.e. all data has been labeled according to its class (0, 1, 2, ..., 9). We can express the category of a digit using a *target vector* $\mathbf{t}$, which represents the identity of the corresponding digit.

Figure 2: Training and test stages

The result of running the machine learning algorithm can be expressed as a function $\mathbf{y}(\mathbf{x})$ which takes a new digit image $\mathbf{x}$ as input and that generates an output vector $\mathbf{y}$, encoded in the same way as the target vectors. Once the model is trained during the *training stage* it can then be used to determine the identity of the new digit images, which are said to comprise the test set. The ability to categorize correctly new examples that differ from those used for training is known as *generalization*.

Pre-processing or Feature Extraction

For most practical applications, the original input variables are *pre-processed* to transform them into some new space of variables where, it is hoped, the pattern recognition problem will be easier to solve. For instance, in the digit recognition problem, the images are typically translated and scaled so that each digit is contained within a box of fixed size. This greatly reduces the variability within each digit class because the location and scale of all the digits are now the same. The pre-processing stage is sometimes called *feature extraction*. Note that new test data must be preprocessed in the same was as the training data.

Figure 3: Feature extraction

Pre-processing might also be performed in order to speed up computation. Here the aim is to find useful features within the data that are fast to compute and yet also preserve useful discriminatory information. These features are then used as inputs to the pattern recognition algorithm. For example in speech processing applications, energy in a few select frequency bands is often used rather than the thousands of sample values present in the waveform of a single word.

Because the number of features is smaller than the number of elements in the data vector, pre-processing represents a form of *dimensionality reduction*. Care must be taken during pre-processing since information is often discarded, and if this information is important for discriminating, system performance may suffer.

Supervised Learning: Classification and Regression

Definition: Applications in which training data comprises examples of the input vectors along with their corresponding target vectors or labels are known as *supervised learning* problems.

Definition: PRML tasks which assign each input vector to one of a finite number of *discrete* categories, are called *classification* problems; if the desired output consists of one of more *continuous* variables, then the task is called *regression*.

The digit recognition problem is an example of a classification problem. An example of a regression problem would be the prediction of the yield in a chemical manufacturing process in which the inputs consist of the concentrations of reactants, temperature, and pressure.

Unsupervised Learning: Clustering, Density Estimation, Visualization

In other pattern recognition problems, the training data consists of a set of input vectors $\mathbf{x}$ without any corresponding target values. The goal in *unsupervised learning* problems may be to discover groups of similar examples within the data, where it is called *clustering*, or to determine the distribution of data within the input space, known as *density estimation*, or to project the data from a high-dimensional space to two or three dimensions for the purpose of *visualization*.

Clustering similar (un-labeled) faces from a collection of photographs or (un-labeled) voices from a collection of audio recordings is an example of unsupervised learning.