CMPSCI 689: Mini Project 2

Due: November 9th 2015

Abstract

This is the second mini project, and covers the material on supervised learning. Read the entire document before starting the project. Your solutions must be computer formatted and submitted online on Moodle by **November 9 2015**. There will be no extensions of this deadline. Each student must turn in an individual solution based on their own work, with no joint work. **Do not wait until the last weekend to work on this project. You will not be able to complete it!**

Dataset

The dataset that you will be analyzing comes from an activity modeling project in the Autonomous Learning Laboratory. Images of people walking up to a B21R robot **Hema** were collected, and processed to extract a variety of features, such as *height* and *width*. The dataset consists of 7 (anonymous) volunteers, and records 15 trials of each subject walking up to the robot. Note that the sequences need not be of the same length! In each file, you will find the height and width data in the first two columns, and the third column is the aspect ratio (height/width). The idea is to try to distinguish the characteristic gait of each person from the rate of change of height and width. The subsequent columns after the first three contain color histogram information. The files `a1` through `a15` contain data about the subject `a`, and so on. The dataset will be put on the class moodle page.

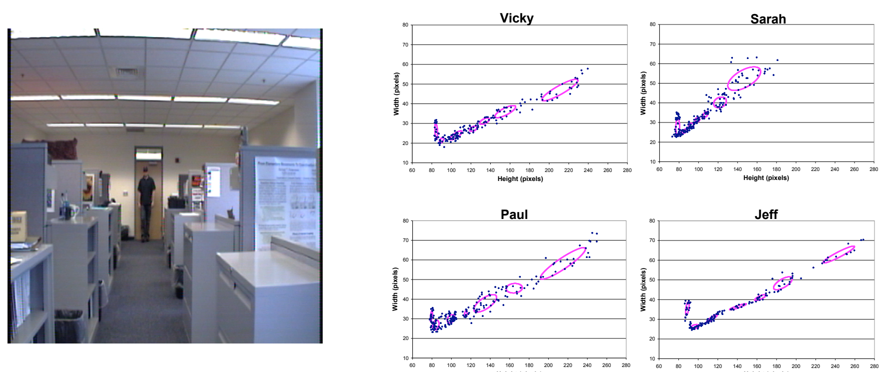

Figure 1: Left and middle: Figure showing how activity data is collected. Right: Sample hidden Markov models trained on the data.

Figure 1 shows an example of the data is shown above, where a student is shown walking up to the robot (seen from the robot's onboard camera). Image processing on the silhouette yields a number of features, such as height, width etc. The goal of the mini project is to train a number of different models and compare their performance. The exact choice of models is left somewhat open, but as a general guide, we suggest using 1) hidden Markov models 2) support vector machines 3) nearest neighbor methods as a somewhat typical comparison. The figure also shows a sample plot of a trained hidden Markov model on four of the data sets.

Theory Question 1 (25 points)

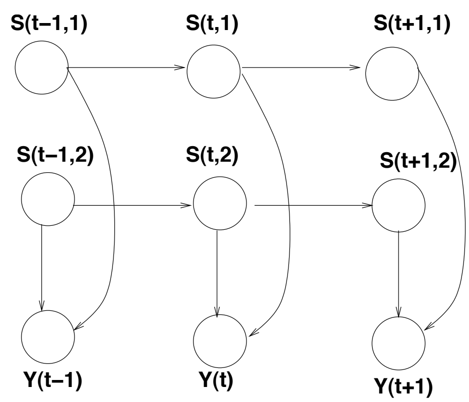

Figure 2: A factorial hidden Markov model.

Figure 2 shows a variant of an HMM, where each state at time t is made up of a vector of state variables $S(t,i)$. Each variable is governed by a probability distribution that depends only on the value of the corresponding variable at the previous time. However, the observation $Y(t)$ at time t can depend on all the state variables at that time.

part a: (12 points) Analyze the conditional independence properties of this variant using the concept of d-separation. In particular, are the state variables at time t marginally independent? Are the state variables at time t conditionally independent given the observation $Y(t)$ at time t? Are the state variables at time t conditionally independent of the past history of state variables, given the value of the variables at the previous time instant $t-1$?

part b: (13 points) Suppose there were M parallel chains (instead of 2 as shown in the figure), and each state variable takes on K values. Show how to convert this variant into a regular HMM, and give an expression for the complexity of the forward algorithm for the converted HMM in terms of T (the length of the sequence), K (the number of values each state variable takes on), and M (the number of chains).

Theory Question 2 (25 points)

The primal form for *kernel ridge regression* can be written as

$$\min_{w,\xi} \lambda| \ w\|^2 + \sum_{i=1}^{l} \xi_i^2$$

$$\text{subject to } y_i - \langle w, x_i \rangle = \xi_i, i = 1, \ldots, l$$

- **part a (10 points):** Derive the Lagrange dual form, as a function of α, y_i, and the inner products $\langle x_i, x_j \rangle$.
- **part b (15 points):** Let G be the Gram matrix for the default linear kernel, that is $(i,j)^{th}$ entry of G is $< x_i, x_j >$. Show that the solution to kernel ridge regression is given by

  $$\alpha = 2\lambda(G + \lambda I)^{-1} y$$

  using which $f(x)$ can be computed as

  $$f(x) = \langle w, x \rangle = y^T (G + \lambda I)^{-1} k$$

  where the vector $k = [\langle x_1, x \rangle, \ldots, \langle x_l, x \rangle]^T$.

Programming Question 3 (25 points)

In this problem, you will train a hidden-Markov model with a mixture of gaussian (MOG) observation model to fit the person trajectories dataset. To simplify your task, we will be giving you MATLAB code that implements the MOG HMM, reads in the data, and outputs the results using the trained model.Read the comments in the code file before you proceed. If you have any problems compiling or running the code, Marwan can help you.

Vary the parameters of the HMM, e.g, the number of states, the number of mixture components etc. and run experiments comparing the performance of the different HMM models. Plot the log likelihood in each case, and compare their classification performance.

Programming Question (25 points)

In this question, you are asked to compare the performance of the HMM above with other classifiers of your choice, principally support vector machines, nearest neighbor methods, and anything else you would like to implement.