INTELLIGENT SYSTEMS QUALIFIER
Spring 2017

Each IS student has two specialty areas. Answer all 3 questions in each of your specialty areas. You will be assigned an identifying number and are required to hand in printed or written copies of your answers with each page identified only by that number. This process enables us to grade your answers anonymously. You should NOT identify yourself explicitly by name on your answer sheets or implicitly by referring in the first person to your work (my project on ABC). Please answer each question starting on a new page, with your answer following the text of the question. Place any relevant references that you cite in an answer at the end of that answer, NOT in a separate section at the end of the whole exam. If you have any questions or feel it necessary to make any assumptions in your answers, do not seek clarification from faculty or staff. Simply record your assumptions as part of your answer.

If one of your areas is **Robotics** , answer the three questions below:

Robotics #1

Path Planning

Path-planning algorithms enable a robot to find a path between two configurations in an unknown, partially known, or known environment. Briefly describe the following four planning algorithms: A*, D*, RRT, and CHOMP, highlighting the differences between them. Summarize your description with a table that lists core distinguishing characteristics of each method.

Which of the above algorithms would you use for a robot operating in a highly dynamic environment with moving obstacles? Justify your choice by describing the benefits of your selected approach and the issues with using the other methods.

Robotics #2

Picking

Imagine a user with a head-up display and wearable computer who is picking parts for a car assembly line for a large manufacturer. In a given 15 minute run, the picker needs to pick 30 parts for each of 6 cars from a warehouse with 200 shelving units in 20 rows. Your job is to create an AI to navigate the picker to the right row and shelving unit and then guide them to picking from the right shelf and bin. The AI must also monitor the picker's hands to check to see that the right bin and correct part is picked from and the part is then placed in the correct (of 6) receiving bins. These bins must then be delivered to the assembly line in time for the next 6 cars to be assembled. The wearable has a wide-angle camera approximately where the user's eyes are, microphones approximately where the user's ears are, and IMUs (accel/gyro/magnetometer) on the user's wrists. The head-up display is much like Google Glass and the speakers can be made loud enough for the industrial environment.

In some senses, the AI guiding the picker is facing the same problems that it would if it were running a robot. Give a list of these problems and the techniques that would be used to address them. How does the problem become harder with a human-in-the-loop? How does it become easier? What other sensors might you will the wearable (or environment) had?

Robotics #3

System Identification (IO-HMMs)

1 Input-Output Hidden Markov Models

Figure 1 shows a typical input-output hidden Markov model (IO-HMM) that describes how a robot can interact with its environment. The environment states x_t are partially observed through the measurements y_t, and the agent's actions u_t directly affect the state. In this graphical model the actions are *exogenous*, meaning they are determined by factors outside of the system.

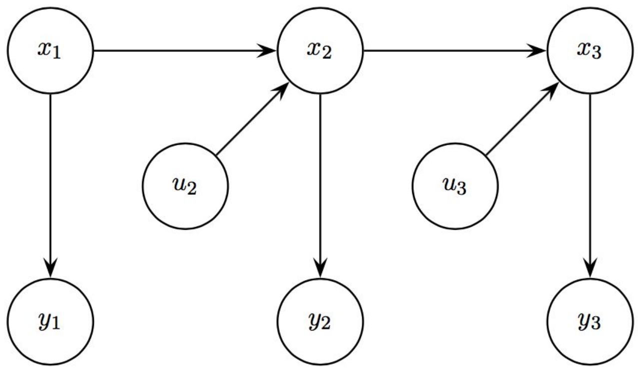


Figure 1: A dynamical system with open-loop control.

How might you learn the parameters of an IO-HMM from dataset D? Assume a "transition model" $p(x_{t+1} \mid x_t, u_{t+1})$, a "measurement model" $p(y_t \mid x_t)$, and a dataset consisting of a length-N sequence of actions and observations, e.g. $\{u_1, y_1, u_2, y_2, \ldots u_N, y_N\}$. First, describe the parameters that must be learned. Then suggest an algorithm that learns these parameters from data and describe how it works. Provide as much detail as you can.

If one of your areas is **Perception**, answer the three questions below:

Perception #1

XKCD

xkcd.com is a comic series with simply drawn "stick figures." However, the artist conveys a sense of personality to each of his figures through simple adornments (see below). While each character is easily identifiable, they may appear in a variety of postures and in different environments. Describe how to create a program which takes the 10 years of xkcd comics and automatically determines the number of unique characters in xkcd and creates a model of each character. Given a new xkcd comic, your method should be able to analyze the picture and identify which of the characters are present. Your method should not depend on knowing the number of characters in advance nor should it use labeled training data. You can and should exploit knowledge of human stick figures. Be specific about what knowledge is learned versus explicitly given to your system, and how your method uses knowledge and images to learn a representation of each character.

Perception #2

Object Detection

This question asks you to contrast two object detection papers from your reading list – Dalal and Triggs, "*Histograms of oriented gradients for human detection*," CVPR 2005 and Girshick, Donahue, Darrell, Malik. "*Region-based Convolutional Networks for Accurate Object Detection and Semantic Segmentation*". PAMI 2015. Both works are highly influential, having 17 thousand and 2.5 thousand citations, respectively. Girshick's RCNN is nearly twice as accurate on the Pascal VOC object detection challenge. How do the systems differ? In particular, what strategies does each method take in terms of (a) search within in image (b) feature representation and (c) training data. What advantages, if any, does Dalal and Triggs have over RCNN?

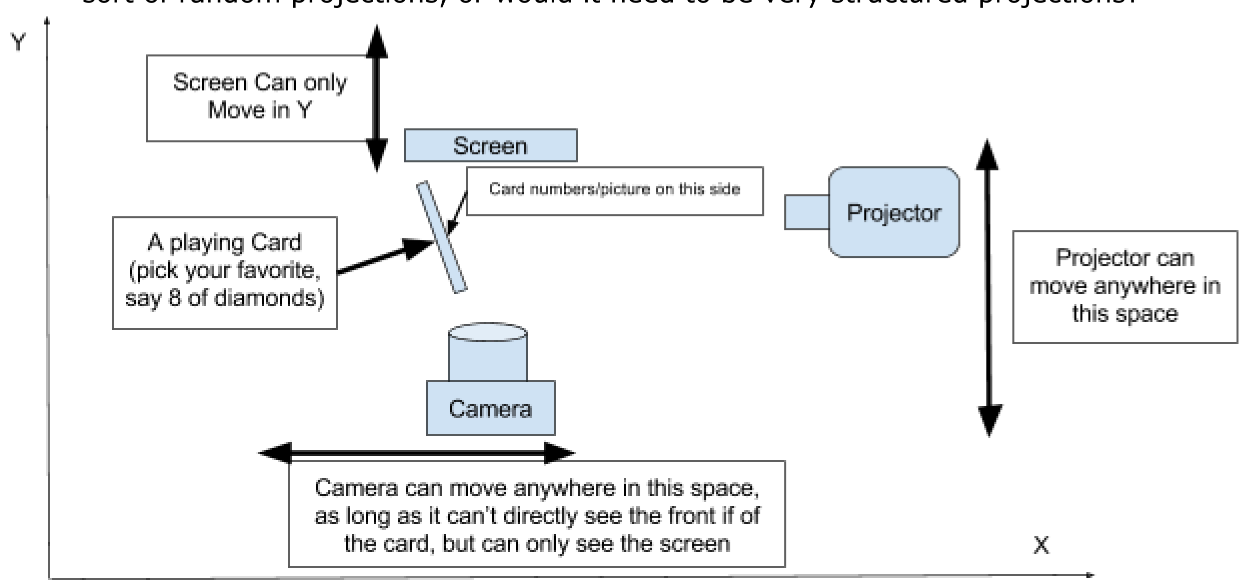
Perception #3

A projector camera system to measure shape and reflectance

You have a camera and a projector connected to a computer. The computer can control each pixel in the projector output independently, so the projector should be viewed as a controllable light source. The camera can capture images. You will use this set-up to accomplish the following tasks. Be sure to state your assumptions for each task clearly.

1. You have been given a statue. How would you generate a 3D model of this statue? Would your methods be different if your projector was
    a. Only Black & White (so can only show binary images),
    b. Grayscale (can only display 256 levels of gray),
    c. Color (can display RGB, each in 256 levels).

    Explain each of the above approaches and state which one would provide the best results, considering that you can't change the resolution of display or the size of the camera sensor. Also, explain your setup briefly (hand drawn diagrams ok, take a picture and include it). Please identify whether each of your methods is an example of a standard computer vision problem/method. Does one need calibration to accomplish this task?
2. Now consider a set-up like the one shown in following diagram. Note the placement of the card, the screen, the camera, and the projector. The playing card is placed such that it CANNOT be seen by the camera and the camera can't be moved so that it can see the card. Using the projector and the camera, do you think it is possible to see what the card is? Feel free to move the camera, the projector and screen. Show the new set-up with figures (hand drawn ok, take a picture and include it). You can also control all other lights in the space. Explain your set-up and your approach. What would your projection be from the projector? Would it be color, grayscale, or black & white? Which would be best? Why? Would you be able to do this by some sort of random projections, or would it need to be very structured projections?



If one of your areas is **Machine learning**, answer the three questions below:

Machine Learning #1

Logistic Regression

Suppose you have two datasets $D_1 = \{x_i^{(1)}, y_i^{(1)}\}_{i \in 1 \dots N_1}$ and $D_2 = \{x_j^{(2)}, y_j^{(2)}\}_{j \in 1 \dots N_2}$, with each $y \in \{-1, 1\}$, and each $x \in \mathbb{R}^P$. The dimension P is identical for D_1 and D_2.

- Let $w^{(1)}$ be the unregularized logistic regression (LR) coefficients from training on dataset D_1, under the model, $P(y \mid x; w) = \sigma(y(x \cdot w))$, with σ indicating the sigmoid function and $x \cdot w$ indicating the dot product of the features x and the coefficients w.
- Let $w^{(2)}$ be the unregularized LR coefficients (same model) from training on dataset D_2.
- Finally, let w^* be the unregularized LR coefficients from training on the combined dataset $D_1 \cup D_2$.

Under these conditions, prove that for any feature n,

$$\begin{aligned} w_n^* &\geq \min(w_n^{(1)}, w_n^{(2)}) \\ w_n^* &\leq \max(w_n^{(1)}, w_n^{(2)}). \end{aligned}$$

Machine Learning #2

Neural Networks / Deep Learning

Neural networks represent a powerful category of statistical classifiers with huge potential for real world pattern recognition and learning tasks. The recent surge in "deep learning" research and especially their numerous practical applications has pushed them (again) at the forefront of both scientific as well as commercial interest.

a) Given the decision theoretical framework of (classification) risk minimisation, specifically the Vapnik-Chervonenkis (VC) dimension / confidence, explain why Neural Networks in general and Deep Learning in particular are favorable for many real-world perception tasks.

For many years the consensus was that neural networks are (almost) universal function approximators thereby requiring only one hidden layer.

b) What was the basis for this consensus and how did it translate into real world applications?

c) How did "deep learning" change this paradigm and why?

DL applications:

d) How can deep neural networks effectively be used for the analysis of time-series data. Discuss in detail at least two different approaches for the analysis of continuous sensor streams from body-worn inertial measurement units (IMUs: sensing platforms that contain 3D accelerometers, gyroscopes and magnetometers -- 9 degree-of-freedom in total, sampling rate 100Hz) within the field of activity recognition.
What are relevant criteria for deciding in favor or against these approaches?

e) Despite the current massive popularity of Deep Learning due to their tremendous success in analysis tasks like object recognition, speech recognition, or natural language processing, fundamental challenges remain that do not allow DL to take off universally.
What is the most pressing current challenge? How do you envision to overcome it?

Machine Learning #3

HMMs and Overfitting

a) List at least three ways an HMM-based recognizer can overfit a training set. Make sure to include situations focusing on each state, each model, and the overall data set.
b) Can state-tying help avoid overfitting?
c) How can state-tying help in situations where not enough training data is available?
d) As an alternative to state-tying one can also optimise the modeling of the feature space (GMMs). Two popular options are codebook sharing (semi-continuous HMMs) and codebook adaptation, e.g., through MLLR or MAP adaptation. Discuss the benefits and challenges of both approaches.
e) One of our recent graduates, Pei Yin, showed that boosting applied to the features of each state of each HMM could be used to significantly improve the recognition rates of a HMM-based recognizer - a technique called Segmentally-Boosted HMMs. Lester and Choudhury, on the other hand, used boosted decision stumps to classify each frame of data coming in and then input those frame-level results to an HMM for final classification. Compare and contrast the benefits, and detriments of these approaches.