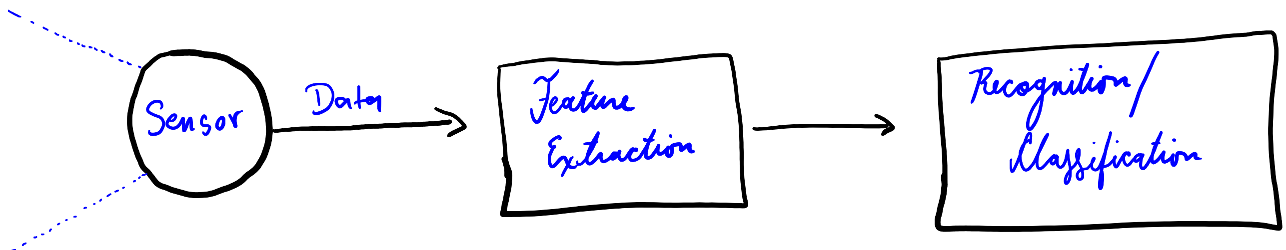
Name: Ravi Patel
PID: A11850926

COGS 118A Spring 2016 - Assignment #1
Due: April 12, 2017, 11:59pm
Grade: ____ out of 100 points
Instructions:
Please answer the questions below and insert your answers to create a pdf file (you can have the hand-written version and scan it to create a pdf file); submit your file to TED (ted.ucsd.edu) by 11:59pm, 04/12/2016. You may search information online but you have to use your own words to answer the questions.
Late policy: 5% of the total points will be deducted for the first late day, then every 10% of the total points will be deducted for every extra day past due.

1 (15 points) Intuition

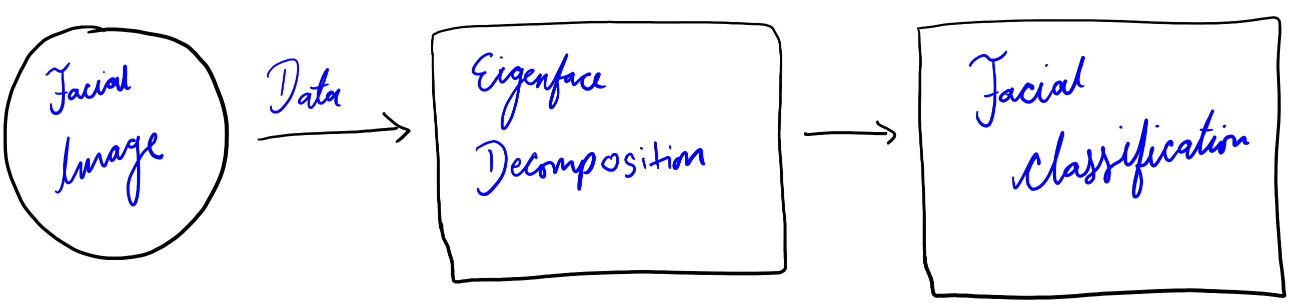
1.1 Please draw a simple diagram illustrating a standard pattern recognition pipeline. Hint: you can copy the one in the slides.



1.2 Please give five examples of real-world classification problems.

1. Facial Recognition/Classification
2. Digit/Handwritting Recognition/Classification
3. Object Recognition/Classification
4. Image Classification
5. Detecting Credit Card Fraud / "Unusual" Pattern Classification

1.3 Please pick one problem from Question 1.2 and design a prototype system to solve it. You may draw a diagram/flowchart to explain your system. For your problem, please clearly describe the input (data) space and output (solution) space.



2 (30 points) IBM Watson

Watch a video clip about IBM Watson at:
https://www.youtube.com/watch?v=3zQI-LMcDnA&t=3s

Instructions: The questions below are supposed to be thought provoking. We will grade your answers based on your general understanding of the problem, as supposed to seeking precisely accurate answers. At this stage, we are just trying to inspire you to think about the problem. We are not going to be very strict. If your answers are valid based on your assumptions, it is all ok. Your answers don't need to be very long. A few lines should suffice.

2.1 Imagine you have unlimited computing resources (for example 100,000 computer cores), how would you design your personal "Watson" to perform the similar tasks as shown in the video? I would make my "Watson" that is capable of predicting the stock market with incredible accuracy.

2.2 What are the possible input data formats of your own Watson? Please write down at least three examples.

- The volume history of each stock
- The price history of each stock
- The divident history of each stock

2.3 Once you have the input data from Question 2.2, how can you turn them into computable representations for IBM Watson as shown in the video?

Each stock is a data vector, where the elements of the vectors are its price, volume, divident, high, low, etc.

2.4 Name three challenges (potential obstacles) to build your own "Watson"?

- Internal influence than can effect stock price
- Changes in the general population behavior
- Natural disaster effecting certain industry (e.g. radiation leaks effecting nuclear industry)

2.5 Please write down three key reasons for IBM Watson's Success. Will they be valid to your Question 2.4 and why?

- Enormous knowledgebase
- Machine learning algorithms
- Learning capability during gameplay (i.e. learn from the correct answer during gameplay)

2.6 Please write down your own explanations about the differences between the following pairs of concepts.

A. Generative vs. Discriminative models
B. Parametric vs. Non-parametric models
C. Supervised vs. Unsupervised learning approaches

A) Generative: learning based on joint probability, $Pr(x,y)$
Discriminative: learning based on conditional probability, $Pr(y|x)$

B) Parametric: based on population parameter & assums population distribution.
Non-Parametric: makes no assumption about population distribution

C) Supervised: knowing the class labels during training
Unsupervised: not knowing the class labels during training

3 (15 points) Event and probability

A fair, six-sided dice is rolled 2 times. Answer the following questions:

Let D_1 & D_2 be the outcome of the first & second roll, respectively. Then D_1 & D_2 are independent.

3.1 What is the probability of the event of rolling an 1 for both rolls?

$$\Pr(D_1 = 1 \cap D_2 = 1) \underset{\text{b/c of independence}}{=} \Pr(D_1 = 1) \cdot \Pr(D_2 = 1) = 1/6 \cdot 1/6 = 1/36$$

3.2 What is the probability of the event of rolling the same number twice?

$$\Pr(D_1 = \alpha \cap D_2 = \alpha) = \sum_{k=1}^{6} \Pr(D_1 = k \cap D_2 = k) \underset{\text{b/c of independence}}{=} \sum_{k=1}^{6} [\Pr(D_1 = k) \cdot \Pr(D_2 = k)] \underset{\text{from 3.1}}{=} \sum_{k=1}^{6} (1/36) = 1/6$$

3.3 What is the probability of the event that the sum of the two rolls add to less than 5?

$$\Pr(D_1 + D_2 < 5) = \sum_{k=1}^{4} \Pr(D_1 + D_2 = k) = \sum_{k=2}^{4} \Pr(D_1 + D_2 = k)$$

$$= \underbrace{\frac{1}{36}}_{k=2} + \underbrace{\frac{2}{36}}_{k=3} + \underbrace{\frac{4}{36}}_{k=4}$$

$$= 7/36$$

4 (10 points) Probability axioms

Probability P is a set function that assigns to sample space Ω, and every subset of it $E, F \subset \Omega$, such that:

1. $P(E) \geq 0$
2. $P(\Omega) = 1$
3. $P(E \cup F) = P(E) + P(F)$ if $E \cap F = \phi$.

With these axioms, one can derive many basic results. Derive the following result: $P(E \cap \bar{F}) = P(E) - P(E \cap F)$.

Recall the conditional probability:

$$\Pr(A|B) = \frac{\Pr(A \cap B)}{\Pr(B)}$$

Recall the Law of Total Probability: $\Pr(A = a) = \sum_{k} \Pr(A = a | B = k) \cdot \Pr(B = k)$.

Pf: By the law of total probability:

$$\Pr(E) = \Pr(E|F) \cdot \Pr(F) + \Pr(E|\bar{F}) \cdot \Pr(\bar{F})$$

by the definition of conditional probability: $\Pr(A|B) \cdot \Pr(B) = \Pr(A \cap B)$

$$\therefore \Pr(E) = \Pr(E \cap F) + \Pr(E \cap \bar{F})$$

Thus, $\Pr(E \cap \bar{F}) = \Pr(E) - \Pr(E \cap F)$.

Q.E.D.

5 (20 points) Matrix

Suppose $\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 5 \end{bmatrix}$, $\mathbf{B} = \begin{bmatrix} 1 & -1 \\ 0 & 2 \\ -1 & 0 \end{bmatrix}$, and $\mathbf{I} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ (***Identity Matrix***), compute

1. $\mathbf{A} + \mathbf{B}$
2. $\mathbf{A}^T\mathbf{B}$
3. $\mathbf{B}^T\mathbf{A}$
4. $\mathbf{B}\mathbf{A}^T\mathbf{I}$

1) $A+B = \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 5 \end{bmatrix} + \begin{bmatrix} 1 & -1 \\ 0 & 2 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 2 & 5 \\ 2 & 5 \end{bmatrix}$

2) $A^TB = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 5 \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 2 \\ -1 & 0 \end{bmatrix} = \begin{bmatrix} -2 & 3 \\ -3 & 4 \end{bmatrix}$

3) $B^TA = \begin{bmatrix} 1 & 0 & -1 \\ -1 & 2 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 5 \end{bmatrix} = \begin{bmatrix} -2 & -3 \\ 3 & 4 \end{bmatrix}$

4) $BA^TI = BA^T = \begin{bmatrix} 1 & -1 \\ 0 & 2 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 5 \end{bmatrix} = \begin{bmatrix} -1 & -1 & -2 \\ 4 & 6 & 10 \\ -1 & -2 & -3 \end{bmatrix}$

6 (10 points) One-hot Encoding

A dataset S is denoted as $S = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5\}$, where each sample refers to the specification of a car. Encode the Make & Color of the car as follows:

Make:

Toyota	1	0	0
Ford	0	1	0
BMW	0	0	1

	Length (inch)	Height (inch)	Make	Color
$\mathbf{x}_1$	186	62	Toyota	Silver
$\mathbf{x}_2$	181	65.5	BMW	Blue
$\mathbf{x}_3$	182.5	59	BMW	Red
$\mathbf{x}_4$	179.5	68	Ford	Blue
$\mathbf{x}_5$	182	53.8	Toyota	Black

Color:

Silver	1	0	0	0
Red	0	1	0	0
Blue	0	0	1	0
Black	0	0	0	1

Please represent this dataset S using a matrix and show it below. **Hint**: for the categorical feature, you may use the one-hot encoding strategy; you can choose either a row vector or a column vector to represent each data sample, but please use one consistently for the entire dataset.

$$S = \begin{bmatrix} 186 & 62 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 181 & 65.5 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 182.5 & 59 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 179.5 & 68 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 182 & 53.8 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}; \text{ i.e. } S = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix}$$