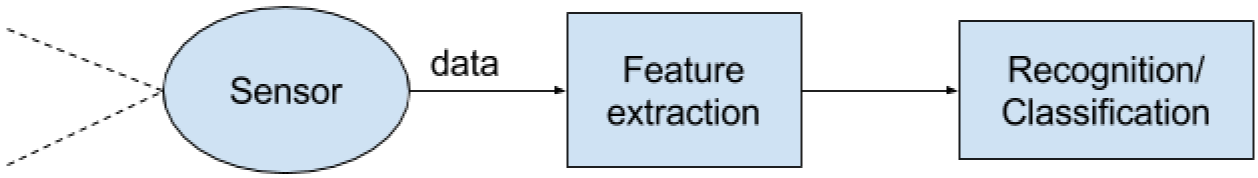
Taylor Tanita
A11216947

COGS 118A: Assignment 1

1 Intuition

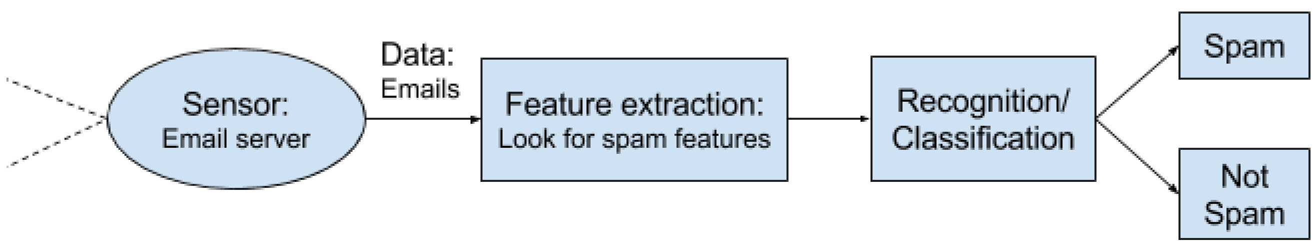
1.1



1.2

1. Email Spam
   Needs to be categorized as spam or not
2. Fraud detection
   Need to determine if a transaction is fraudulent or not
3. Sound recognition
   Recognize what words are being said based on sound input
4. Facial recognition
   Recognize what faces belong to what person
5. OCR- character recognition
   Recognize what character is written

1.3



Input space: email inbox
Output space: different email folders

Taylor Tanita
A11216947

2 IBM Watson

2.1. Imagine you have unlimited computing resources, how would you design your personal "Watson" to perform the similar tasks as shown in the video?

- I would design my own personal Watson to be able to tell me what food would most satisfy me at the time. With so many options, especially in San Diego, I often struggle to figure out what I want to eat most. If I were able to have Watson calculate things such as my mood, hunger level, the time and occasion, it would be great if he could tell me where or what I should eat.

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

1. What I eat everyday: what kinds of food I eat and at what times
2. Where I eat everyday: The locations of where I eat- I am more likely to be satisfied going to places that are similar to where I usually go.
3. Who I eat with, for what occasion: If it is a brunch day, my meal is much different from a normal morning on the go. Additionally, I tend to eat at the same places with the same places; for example, I always go to Urban Plates with my parents but go to World Curry with my friends.
4. Types of food available around me- at grocery stores, restaurants, in my house.

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?

- I would have to use a decent amount of one-hot encoding to encode food items, locations, occasion. However, I could use number representations for things like time and hunger level.

2.4. Name three challenges to build you own "Watson"

1. I go through phases- sometimes I eat burritos a lot, but after a while I can get tired of them. If I have been eating burritos a lot lately, Watson would probably suggest burritos- however, I could not want to eat them anymore.
2. I could not know my own hunger level- sometimes I think I am more hungry than I actually am where as others, I don't think I am as hungry as I am. Watson will have to recognize the occasions where I could be wrong.
3. Watson knowing the significance of the people I am eating with- if I am eating with someone new, Watson won't know if it a friend, a business meeting, or a date.

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

1. The immense amount of data inputted: I could input extensive patterns of my eating as well as food around me. My Watson could succeed with the amount of data I provide.

Taylor Tanita
A11216947

2. Watson's ability to learn on the fly- for example when the answers to that one section in jeopardy were all months. I believe Watson could eventually learn to distinguish the occasion of my eating based on who I am eating with.
3. Watson's ability to see through tricky phrasing like puns- my Watson could see maybe if a food meeting with a friend has other intentions- such as work connections or romance.

2.6. Explanations about differences

A. Generative vs. Discriminative models
   Generative can constructs models; it can reconstruct data. Discriminative performs classification; it can differentiate what kind of data it is. A simple example would be being able to recite the digits of pi vs recognizing the digits of pi. The recitation is generative because you are reproducing the values of pi- this is rare in many humans beyond the first few digits. The recognition that a number is pi on the other hand- if fairly common knowledge. This is discriminative.
B. Parametric vs. Non-parametric models
   Parametric has a formula with defined input(s) and output- the complexity of the formula is constant. Non-parametric has many inputs that must be considered and as the number of inputs increase, the output becomes more complicated. An example of this would be if you want to know if the mall will be crowded. In a parametric model, you could consider the time of year, time of day, and what sales are going on. In a non-parametric model, you would consider the crowd at the mall every day in its history.
C. Supervised vs. Unsupervised learning approaches
   In supervised learning approaches, we have all the data and all the labels during training. In unsupervised learning approaches, the data does not include labels. For example, if you wanted to be able to predict the age at which someone will get married, in a supervised learning approach, you would train the program with information on the age they got married. In an unsupervised approach, you would not include the age- the computer would have to cluster appropriately to predict this without the knowledge of when the people actually got married.

Taylor Tanita
A11216947

3 Event and probability

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

- $\frac{1}{6} * \frac{1}{6} = \frac{1}{36}$

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

- $6\,(\frac{1}{6} * \frac{1}{6}) = \frac{1}{6}$

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

- Possible valid rolls: (1,1), (1,2), (1,3), (2,2), (2,1), (3,1) = 6
- $\frac{6}{36} = \frac{1}{6}$

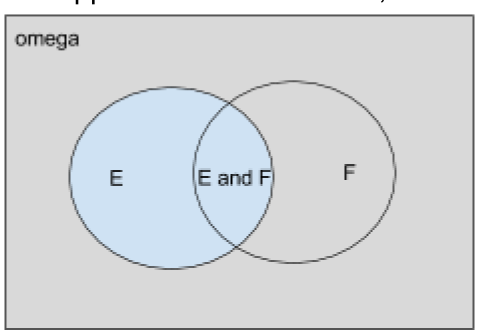
4 Probability axioms

Probability P is a set of functions 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. $P(E \cup F) = P(E) + P(F)\ if\ E \cap F = \emptyset$

Derive $P(E \cap \overline{F}) = P(E) - P(E \cap F)$

- To supplement the derivation, a venn-diagram is provided that will be referred to



- By adding $P(E \cap F)$ to both sides, we get $P(E \cap \overline{F}) + P(E \cap F) = P(E)$
- From the diagram, we can see that $E \cap F$ and $E \cap \overline{F}$ are disjoint. Thus if we apply rule 3 to this and let $E = E \cap F$ and $F = E \cap \overline{F}$, we can the values into rule 3 $P((E \cap F) \cup (E \cap \overline{F})) = P(E \cap F) + P(E \cap \overline{F})$
- $E = (E \cap F) \cup (E \cap \overline{F}),\ \emptyset = (E \cap F) \cup (E \cap \overline{F})$
- $P((E \cap F) \cup (E \cap \overline{F})) = P(E)$
- Thus $P(E \cap \overline{F}) = P(E) - P(E \cap F)$

5. Matrix

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 3 \\ 3 & 5 \end{pmatrix} \qquad B = \begin{pmatrix} 1 & -1 \\ 0 & 2 \\ -1 & 0 \end{pmatrix} \qquad I = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

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. $A^T B$

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

3. $B^T A$

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

4. $B A^T I$

$$\begin{bmatrix} 1 & -1 \\ 0 & 2 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 5 \end{bmatrix} = \begin{bmatrix} -1 & -1 & -2 \\ 4 & 6 & 10 \\ -1 & -2 & -3 \end{bmatrix}$$

6. One-hot Encoding

Silver	1000	Toyota	100
Blue	0100	BMW	010
Red	0010	Ford	001
Black	0001		

$$\begin{bmatrix} 186 & 62 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 181 & 65.5 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 152.5 & 59 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 179.5 & 68 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 182 & 53.8 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$