

Homework 2

Due: 19 Sep 2017

REMEMBER:

- Include work in symbolic form (e.g. $p(X = \text{foo} | Y = 3)$). I should be able to tell where every number came from. You can abbreviate (e.g. $p(\text{foo} | 3)$) as long as it's clear what is meant.
- Sanity-check your answers. If your answer seems crazy but you can't find the error, at least make it clear you know there's a problem.

Problem 2.1

Recapping from the previous homework: A standard deck of playing cards has 52 cards in four suits, each suit with cards numbered 2–10 and “face cards” labelled jack, queen, king, and ace. A “pinochle deck” has 48 cards: each of the four suits has only 9 and 10 plus the face cards, and each card appears twice in the deck.

I have many decks of playing cards, most of which are standard, but some are pinochle decks—for the purposes of this problem, let's say I have two pinochle decks and ten standard decks. Without counting the cards, it's hard to tell at a glance whether you've accidentally grabbed a pinochle deck.

- a. If I grab a deck completely at random and draw a card from it, what is the probability that the card is a 5?
- b. If I grab a deck completely at random and draw a card from it, what is the probability that the card is a jack?
- c. If I grab a deck completely at random and draw a card, and the card is a jack, what is the likelihood that I've grabbed a pinochle deck?
- d. If I grab a deck completely at random and draw two cards from it, some pairs give me certainty: if either card is a 4, for instance, or if both cards are the jack of diamonds. But if one is a jack and the other is a king, does that give me any knowledge about the deck? Why or why not? (Note: I'm not looking for exact numbers on this part, because they're subtle and a bit gross. Focus on the analysis.)

Again, don't forget to show your work.

Problem 2.2

You are evaluating (by hand) a piece of image recognition software that attempts to retrieve images of individual people from an image base and label them as male or female.

The evaluation corpus has 180 images in it. Of these, 60 are landscapes, 20 are crowded urbanscapes, 20 mugshots, 50 portraits, and 30 candid photos of individuals. The mugshots, portraits, and candid are each equally divided into pictures of men and pictures of women. The system yielded the following results:

Actually a:	Guessed it was:	
	male	female
Landscape	2	1
Urbanscape	2	3
Male Mugshot	8	2
Female Mugshot	3	7
Male Portrait	21	2
Female Portrait	3	21
Male Candid	6	1
Female Candid	0	8

- First, evaluate the simpler task of retrieving the images of individuals (either male or female, even if the system gets that piece wrong). What is the system's precision? Recall?
- How does the evaluation change when you take into account the label the system places on extracted images—that is, incorrectly labelled images are simply incorrect?
- What uses might a system like this be put to?
- Set aside simple precision and recall for a moment and look at the actual results—are there some subtasks it is better at? How does the composition of the test corpus affect the results? How good is this test corpus and this evaluation at actually evaluating the system for its intended uses?