BBL588E Final Takehome Solutions

Yüksel Kapan

704191018

July 2, 2020

1 Question 1

You may use your favorite programming environment/language/tool for the following question

- Download the İTÜ Lake – Bee image.

- Design a method, possibly composed of several sub-methods, to segment out the bee in the middle of the image.

- You may use a combination of high-/low-pass filters, segmentation methods, and/or classifiers. These are called hand-crafted feature extraction based algorithms.

- Instead of designing your own segmentation algorithm, you are free to use convolutional neural network (CNN) based segmentation algorithms, such as SegNet, Mask R-CNN, as well.

- You may even use a combination of CNN based techniques along with hand-crafted feature extraction based algorithms.

- Measure the performance of your method. For that purpose, you first need to generate the ground-truth label by using an image annotation tool. A list of tools is provided below. You may use one of them to label "bee".

- Then, use intersection-over-union as a metric to measure the performance of your method.

For manual image segmentation, I used an open source project called Via. The link to the project is: https://gitlab.com/vgg/via

After segmenting the image and creating the label, I first converted the image into HSV color space. Then, I tried a bunch of threshold values for Hue and Saturation, and found a good combination of thresholds that distinctively separated the bee from the background. Then I ran edge detection algorithms and in the detected items list, I found out the index of the bee. With the bee index, I was then able to calculate the IoU score.

The source code and the results can be found at: https://github.com/yukosgiti/BBL588E$_{fin}$

2 Question 2

Provide your solutions for the following Bayes Decision Rule examples.

Example 1 You go to see the doctor about dry coughs and fever you have recently. The doctor directs you to have a PCR test for COVID-19. Assume COVID-19 affects 2167 people in 1,000,000 people in Turkey. Assume also that the test is 99.5% accurate, in the sense that the probability of a false positive is 0.5%. The probability of a false negative is 20%. You test positive. What is the probability that you "actually" have COVID-19?

Solution 1 There a two states in which a person is classified. One is whether they are *Diseased* or *Healthy* and the other one is whether the test result for them is *Positive* or *Negative*.

Let's put the possible states in a table for better visualization.

	Diseased	Healthy
Test is (+)	TP	FP
Test is (−)	FN	TN

Figure 1: The table describing the possible states.

Now that we have defined our variables, let's see what the question asks us in terms of these variables.

$$P(Diseased|Positive) = \frac{TP}{TP + FP} = ?$$

Let's also redefine the information in the question in terms of our new variables.

$$P(Diseased) = \frac{TP + FN}{TP + FP + FN + TN} = 2167/10^6$$

$$P(FalsePositive) = \frac{FP}{FP + TN} = 5/10^2$$

$$P(FalseNegative) = \frac{FN}{FN + TP} = 20/10^2$$

From these information we can also infer the following:

$$P(Healthy) = \frac{FP + TN}{TP + FP + FN + TN} = 1 - P(Diseased)$$

$$P(TrueNegative) = \frac{TN}{FP + TN} = 1 - P(FalsePositive)$$

$$P(TruePositive) = \frac{TP}{FN + TP} = 1 - P(FalseNegative)$$

Now that we have the necessary information we can use the Bayes' Rule to solve the question.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Figure 2: Bayes' Rule.

Remember the question was asking $P(Diseased|Positive)$. Let's apply the Bayes' Rule.

$$P(Diseased|Positive) = \frac{P(Positive|Diseased)P(Diseased)}{P(Positive)}$$

First let's find $P(Positive|Diseased)$. This asks us the probability of test result being positive given a person is actually diseased. From Figure **??** we can see that this corresponds to $TP/(TP + FN)$ which is actually $P(TruePositive)$.

We already know what $P(Diseased)$ is. Therefore, only remaining unknown is $P(Positive)$.

Again from Figure **??** we can see that $P(Positive)$ is $(TP + FP)/(TP + FP + FN + TN)$. Which is something that we don't know the answer yet. However it is something we can derive with the information we have by using some algebra.

$$\frac{TP + FP}{TP + FP + FN + TN} =$$

$$\frac{TP + FN}{TP + FP + FN + TN} \cdot \frac{TP}{TP + FN} + \frac{FP + TN}{TP + FP + FN + TN} \cdot \frac{FP}{FP + TN}$$

$$= P(Diseased)P(TruePositive) + P(Healthy)P(FalsePositive)$$

4

Now let's rewrite $P(Diseased|Positive)$ with the things we know.

$$= \frac{P(TruePositive)P(Diseased)}{P(Diseased)P(TruePositive) + P(Healthy)P(FalsePositive)}$$

Example 2 Now imagine that you went to a friend's wedding in Spain recently (although it is forbidden to travel, for the sake of this very question assume you did so), and it is know that 6240 in 1,000,000 people who visited Spain recently come back with COVID-19. Given the same test result as above, what should your revised estimate be for the probability you have the disease?

Solution 2 Deneme