Assignment 2
Decisions, decisions, decisions!

Homeworks Guidelines and Policies

- ***What you must hand in.*** It is expected that the students submit an assignment report (HW2_[student_id].pdf) as well as required source codes (.m or .py) into an archive file (HW2_[student_id].zip).
- ***Pay attention to problem types.*** Some problems are required to be solved *by hand* (shown by the icon), and some need to be implemented (shown by the icon).
  Please don't use implementation tools when it is asked to solve the problem by hand, otherwise you'll be penalized and lose some points.
- ***Don't bother typing!*** You are free to solve by-hand problems on a paper and include picture of them in your report. Here, cleanness and readability are of high importance. Images should also have appropriate quality.
- ***Reports are critical.*** Your work will be evaluated mostly by the quality of your report. Don't forget to explain what you have done, and provide enough discussions when it's needed.
- ***Appearance matters!*** In each homework, 5 points (out of a possible 100) belongs to compactness, expressiveness and neatness of your report and codes.
- ***Python is also allowable.*** By default, we assume you implement your codes in MATLAB. If you're using Python, you have to use equivalent functions when it is asked to use specific MATLAB functions.
- ***Be neat and tidy!*** Your codes must be separated for each question, and for each part. For example, you have to create a separate .m file for part b. of question 3. Please name it like p3b.m.
- ***Use bonus points to improve your score.*** Problems with bonus points are marked by the icon. These problems usually include uncovered related topics or those that are only mentioned briefly in the class.
- ***Moodle access is essential.*** Make sure you have access to Moodle because that's where all assignments as well as course announcements are posted on. Homework submissions are also done through Moodle.

- ***Assignment Deadline.*** Please submit your work **before the end of December 26th**.
- ***Delay policy.*** During the semester, students are given 7 free late days which they can use them in their own ways. Afterwards there will be a 25% penalty for every late day, and no more than three late days will be accepted.
- ***Collaboration policy.*** We encourage students to work together, share their findings and utilize all the resources available. However you are not allowed to share codes/answers or use works from the past semesters. Violators will receive a zero for that particular problem.
- ***Any questions?*** If there is any question, please don't hesitate to contact me through **ali.the.special@gmail.com**.

1. Do You Think You May Have COVID-19? (10 Pts.)

Keywords: *Bayesian Inference (Reasoning), Prior Probability, Posterior Probability, Likelihood Function*

To date, COVID-19 has infected almost 69 million people and killed over 1.5 million around the world. Despite promising news about the production of Coronavirus vaccines and the beginning of mass inoculation in Britain, the increasing number of patients in populated countries including USA, Brazil and India has caused serious concerns.

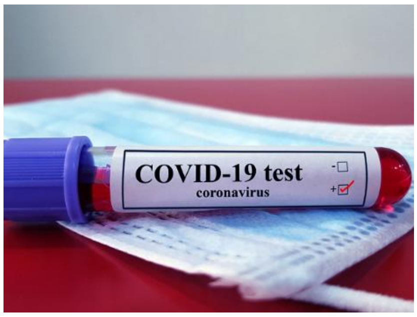


Figure 1 COVID-19 tests are of different degree of reliability, ranging from 68% to 100% in sensitivity.

Several methods have been introduced to assess the current or past presence of COVID-19 virus in body, with some more reliable than the others. In this problem, we're going to use **Bayes Theorem** and analyse the results obtained from different tests in different occasions.

First, assume you have been told that you tested positive for COVID-19. However, you live in Philippines, where the virus is relatively rare and only four in a thousand people have been infected. On the other hand, the sensitivity of the test is 83% and the false positive and false negative rate are 17% and 11%, respectively.

a. How worrying is the situation? Use Bayes' theorem to justify your answer.

In a different scenario, you are in Australia, where the virus has infected only 1 out of 1,000 population. The test is 73% accurate, and the probability of a false positive and false negative are 27% and 19%, respectively. You test positive.

b. What is the probability that you actually have COVID-19?

However, you suddenly remember that you've just came back from Belgium, where 50 out of 1,000 people had been infected.

c. How does your answer in the previous part change?

Next, you are a doctor and suspect that one of your patients may have been infected with COVID-19. You order a CRP test, in which 76% of the results are positive when the patient actually have the virus, and 68% of the results are negative when the patient doesn't have the virus. Also, you work in the UK, where out of 1,000 people, 25 have tested positive.

d. How likely is your patient to actually have coronavirus when he tests negative?
e. How about if the results show positive?
f. If your patient tests positive, is he more likely to actually have Coronavirus than not?
g. If your patient tests negative, is he more likely to not have Coronavirus than have it?

Finally, you are in Turkey, where 10 out of 1,000 people have Coronavirus. There is a new test with the probability p of being correct. In other words, p of those with Coronavirus test positive and $1-p$ of those without also test positive.

h. If a patient test positive, what is the probability of him actually has Coronavirus?
i. Let q be the probability that a patient actually has Coronavirus if the test is positive. For what numerical value of q do you consider the test as being reliable?
Hint: There is no "correct" answer. Just choose any number between 0 and 1.
j. Based on parts h. and i., what should p be for you to consider the test as being reliable?

2. Making Decisions using Bayes Decision Rule (12 Pts.)

Keywords: *Classification Problem, Bayes Decision Rule*

Bayes Decision Rule is a decision theory which is informed by **Bayesian Probability**. By using probabilities and costs, Bayes decision rule tries to quantify the trade-off between various decisions. A classifier who applies such a decision theory uses the concepts of Bayesian statistics to estimate the expected value of its decisions.

In this problem, X is a two dimensional feature vector and $p(\mathrm{x} \mid \omega_i) \sim N(\mu_i, \Sigma_i)$, $i = 1, 2$, as its class conditional distribution where

$$\mu_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \qquad \mu_2 = \begin{bmatrix} -1 \\ -3 \end{bmatrix}$$

and

$$\Sigma_1 = \begin{bmatrix} 1.8 & -0.7 \\ -0.7 & 1.8 \end{bmatrix} \qquad \Sigma_2 = \begin{bmatrix} 1.5 & 0.3 \\ 0.3 & 1.5 \end{bmatrix}$$

a. Sketch the contours of constant values for two class conditional densities.
b. Sketch the decision boundary for Bayesian classifier and minimum distance classifier.
c. Generate 1000 samples for each class and estimate the classification error for each classifier.
d. Considering the following cost matrix, find the decision boundary for Bayes classifier and compare f-score for generated samples with Bayes classifier in part b.

$$\begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix} = \begin{bmatrix} 0 & 3 \\ 1 & 0 \end{bmatrix}$$

e. Repeat part c. for 20 times, and report an average error for both classifiers.

Next, consider 6 different datasets given in Figure 2. The goal is to design a Bayes classifier assuming Gaussian distribution for the data. The covariance matrices are not equal across classes, but they are diagonal on the form $\Sigma_i = \sigma^2 \mathbf{I}$.

f. In each case, sketch the Bayes decision boundary. If the classifier breaks down, explain.

Hint: You can use the provided images of these datasets (attached to this homework) and an image editor software (like Microsoft Paint).

Figure 2 Six different datasets with different distributions

3. Automatic Saffron Cleaning Machine (18+6 Pts.)

Keywords: *Classification Problem, Bayes Decision Rule, Minimum Distance Classifier, Confusion Matrix, Bayes Error, ROC Curve*

Saffron is the world's most expensive spice and one of the most valuable substances in the world. High-quality saffron can cost you over $10,000 per kilogram in Europe and North America. This spice is so precious because the harvesting process is labour-intensive. Each flower produces just a few stigmas, which required to be hand-picked and dried before being used with food.

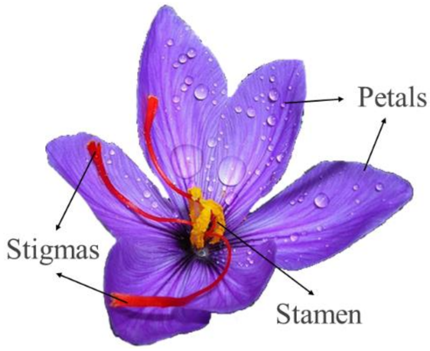


Figure 3 Saffron flower consists of three main parts, from which the red part (stigma) is separated and used in further processes.

In this problem, we aim to design a Bayesian classifier capable of separating saffron flower into its three main parts ("petal", "stamen" and "stigma") using image frames recorded by a camera. We address this problem by taking into consideration different set of features extracted from saffron flower image. The feature sets are in different dimensions and their details are as below:

- *{Y/B Tones}* which is the third channel of the image in LAB color space and denotes the amount of yellow or blue tones in the image. A large positive value corresponds to yellow and a large negative value corresponds to blue.
- *{X-Derivative, Y-Derivative}* which are obtained from applying horizontal and vertical derivative filters on the grayscale image. Large values indicate sudden horizontal or vertical changes (i.e. edges) in the image.
- *{Hue, Saturation, Value}* which are different channels of the image in HSV color space.

Table below shows the corresponding values for each of these features gathered from 15 different pixels of a sample flower image, as well as the region they belong to (i.e. their labels).

#	Label	Feature Set 1	Feature Set 2		Feature Set 3		
		Y/B Tones	X-Derivative	Y-Derivative	Hue	Saturation	Value
							
1	Petal	-47.55	1	0	0.73	0.73	0.51
2	Petal	-70.45	29	0	0.74	0.73	0.88
3	Petal	-55.91	0	0	0.73	0.51	0.95
4	Petal	-12.55	0	0	0.78	0.64	0.14
5	Petal	-57.28	28	0	0.73	0.50	1.00
6	Stamen	74.56	0	14	0.13	0.91	0.86
7	Stamen	71.91	0	0	0.12	0.97	0.85
8	Stamen	25.57	5	0	0.03	0.58	0.65
9	Stamen	33.19	8	12	0.05	0.61	0.75
10	Stamen	79.07	0	6	0.13	1.00	0.93
11	Stigma	44.35	7	0	0.98	0.97	0.77
12	Stigma	27.83	96	0	0.01	0.54	1.00
13	Stigma	47.32	0	0	0.01	0.88	0.77
14	Stigma	54.10	6	0	0.02	0.85	0.94
15	Stigma	56.82	1	5	0.00	0.88	0.97

a. Design three Bayesian classifiers using each of the feature sets. The data are assumed to have Gaussian distribution with the same covariance matrix $\sum = \mathbf{I}$. For each of the classifiers, find the general form of the discriminant function.

b. A new saffron flower has arrived and an image has been recorded. The following features has been extracted from different parts of the image. Use the discriminant functions obtained in the previous part and predict their regions using each of the three classifiers.

#	Feature Set 1	Feature Set 2		Feature Set 3			True Label
	Y/B Tones	X-Derivative	Y-Derivative	Hue	Saturation	Value	
1	54.73	0	3	0.07	0.98	0.71	Stamen
2	51.08	6	4	0.01	1.00	0.69	Stigma
3	-44.32	3	8	0.73	0.40	0.98	Petal
4	53.91	0	3	0.00	0.97	0.78	Stigma
5	-66.50	0	5	0.75	0.67	0.91	Petal
6	68.64	0	17	0.10	0.93	0.87	Stamen
7	-62.97	0	7	0.74	0.71	0.78	Petal
8	56.47	0	6	0.07	1.00	0.71	Stamen

c. Compute a confusion matrix for each classifier.

d. Calculate the Bayes error for each classifier.

e. Draw a ROC curve to visualize the performance of each classification.

f. Repeat the previous parts with a MDC classifier. Display and compare the results.

Note: Using MATLAB or similar tools is allowed (and recommended) for calculations. However, the whole process as well as final results must be included in the report.

4. Dealing with Prediction Error in Bayes Decision Rule (10 Pts.)

Keywords: *Bayes Decision Rule, Probability of Error, Upper Bounds of Error Probability, Bhattacharyya Error Bound, Chernoff Error Bound*

In general, the **Bayes Decision Rule** – or any other decision rule – does not lead to perfect classification. In order to measure the performance of a decision rule, one must calculate the **Probability of Error**, which is the probability that a sample is assigned to a wrong class.

In practice, calculating the error probability is a difficult task. We may seek either an approximate expression for the error probability, or an upper bound on the error probability. **Bhattacharyya Error Bound** and **Chernoff Error Bound** are some **Upper Bounds of Error Probability**.

Assume X is a one dimensional feature which is used to decide between two categories ω_0 and ω_1, with conditional densities as depicted in Figure 4.

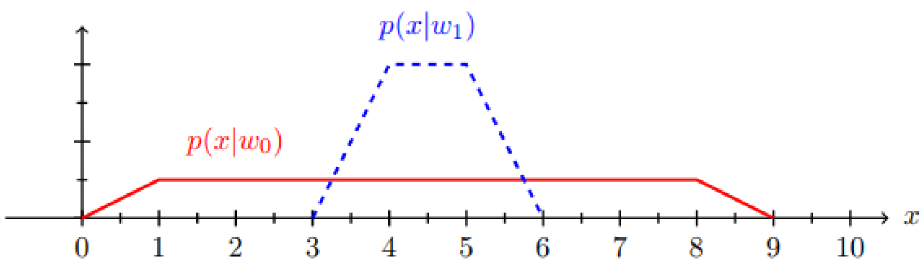


Figure 4 Class-conditional probability distributions

Assuming $p(\omega_0) = 0.3 p(\omega_1)$,

a. Determine and sketch the decision regions for a Bayes decision rule.
b. Compute the probability of error for this rule.
c. Calculate the Bhattacharyya error bound.
d. Calculate the Chernoff error bound.
e. Determine the Neyman-Pearson classifier as well as its error. Assume $\varepsilon_1 = 0.05$.

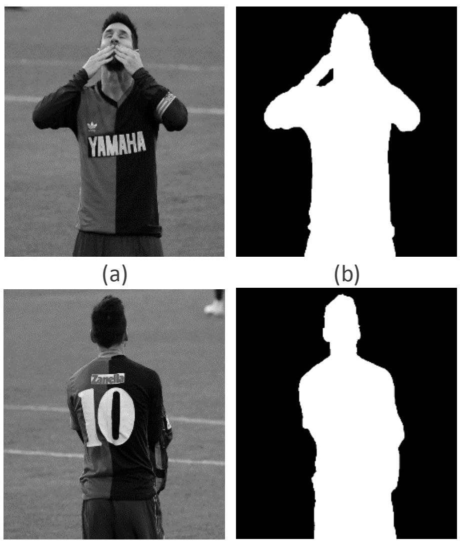
5. Separating Messi from Football Field (15 Pts.)

Keywords: *Classification Problem, Bayes Decision Rule, Image Segmentation, Discrete Cosine Transform*

(a) (b)

Figure 5 A simple example of image segmentation, where the image is divided into two regions (a) original image (b) segmented image

So far, you've seen several problems in which the theory of Bayes decision rule has been addressed. Now, we're about to be amazed by the performance of this method in a real-world computer vision task; **Image Segmentation**, where the goal is to divide a digital image into multiple segments in order to simplify it for further manipulations. For the sake of simplicity, here we only consider classifying an image into two parts, "Messi" (foreground) and "Field" (background). We also consider images in grayscale.



(a) (b) (c) (d)

Figure 6 Train and test images alongside their masks (a) train image (b) train mask (c) test image (d) test mask, which is used for evaluation purposes

In order to define this task in a pattern recognition problem framework, we must first choose which features to select. To do so, we first consider space of 8×8 image blocks, i.e. images are perceived as collections of 8×8 blocks. Then for each block, the discrete cosine transform is computed and a matrix of 8×8 frequency coefficients will be obtained. The intuition behind this is that "Messi" and "Field" have distinctive textures with different frequency decompositions, hence the two classes are expected to be better separated in the frequency domain. We further convert each block into a vector using a zigzag pattern, and assume the index of the coefficient that has the second largest absolute value as a feature representing that block, which is denoted by X. We finally build a histogram of these indexes to obtain class-conditionals for two classes $P_{X|Y}(x\,|\,messi)$ and $P_{X|Y}(x\,|\,field)$. Another similar image is also provided in order to find the values of priors $P_Y(messi)$ and $P_Y(field)$.

a. Load the image "leo1.png" as the training image.
   Perform the above procedure and find reasonable estimates for the prior probabilities.
b. Use the training image to compute and plot the index histograms $P_{X|Y}(x\,|\,messi)$ and $P_{X|Y}(x\,|\,field)$.

c. Now load the image "leo2.png" as the test image. For each block in this image, calculate the feature X .
d. Compute the state variable Y using the minimum probability of error rules based on the probabilities you calculated in the previous parts. Store the state in a matrix and display the result as an image.
e. Use the ground truth mask provided for the test image and compute the probability of error.

Recommended MATLAB functions: `imread()`, `im2double()`, `dct2()`, `imagesc()`, `colormap()`

6. Maximum Likelihood Approach for Parameter Estimation (10+4 Pts.)

Keywords: *Parameter Estimation, Maximum Likelihood Estimation, Probability Mass Function, Sufficient Statistics*

Maximum Likelihood Estimation (**MLE**) is a **Parameter Estimation** method which tries to find the parameter values of a statistical model that maximise the **Likelihood Function**, given the observations. The resultant is called **Maximum Likelihood Estimate**, abbreviated as **MLE**.

Let X be a discrete random variable with the following probability mass function, with parameter θ where $0 \le \theta \le 1$.

The following 10 independent observations were generated from this distribution: $(1,2,1,3,4,2,3,1,1,3)$.

X	*P(X)*
1	$3\theta/5$
2	$2(1-\theta)/5$
3	$2\theta/5$
4	$3(1-\theta)/5$

a. What is the likelihood function $L(\theta)$?
b. Find the log likelihood function.
c. Using one of the above functions, determine the maximum likelihood estimate of θ .

Now, assume $x_1, x_2, ..., x_n$ be i.i.d. samples from *Erlang* distribution, with unknown parameter θ :

$$f(x \mid \theta) = \frac{1}{(m-1)!}\left(\frac{1}{\theta}\right)^m x^{m-1} e^{-\frac{x}{\theta}}, \quad x \ge 0, \quad \theta > 0$$

d. Find the maximum likelihood estimate of the parameter θ .

Finally, suppose that n samples $y_1, y_2, ..., y_n$ are drawn independently from a continuous distribution given by:

$$f(y_i, \theta) = \frac{e^{-\theta}\theta^{y_i}}{y_i!}$$

e. Find $\hat{\theta}_{ML}$.
f. Prove that $\hat{\theta}_{ML}$ is unbiased for θ and find its variance.
g. Assuming the gamma distribution as the prior distribution for θ, compute the posterior distribution and posterior mean of θ. Compare ML and MAP estimates.
h. Show that $T(y) = \sum_{n=1}^{N} y_n$ is a sufficient statistic for θ .
i. Find the **Cramer-Rao Lower Bound** on the variance of the unbiased estimators of θ. Is $\hat{\theta}_{ML}$ efficient? Explain.

7. Here's Why Hitler Hated MLE So Bad (12 Pts.)

Keywords: *Parameter Estimation, Maximum Likelihood Estimation, Biased Estimator, Estimator Variance, German Tank Problem*

German Tank Problem is one of the most well-known **MLE** problems. In World War II, the Allies attempted to determine the total number of German tanks, known as *Deutsche Panzer*, by looking at the serial numbers of those they had captured or destroyed.

Figure 7 A German Panzer in the streets of Paris. Allied technicians applied statistical approaches to estimate the number of German tanks during World War II.

Assume you are a British engineer who works for the UK army, and you are trying to give an estimate of the total number of German tanks to your commanding officer. The available data are the number of spotted tanks k and the highest spotted serial number m. You also assume that the tanks were labelled sequentially from 1 to n, and that the serial numbers are randomly and uniformly distributed among the tanks in the battlefield.

a. Assuming n tanks are in the battlefield and k have been observed, find the likelihood of the highest observed serial number being m.
b. Find the posterior $P(n|m,k)$, considering a flat prior between the number of spotted tanks k and some maximum number Ω such that $P(n)=\begin{cases}1/\Omega & \text{for } 1\le n<\Omega \\ 0 & \text{otherwise}\end{cases}$. Take the limit $\Omega\to\infty$.

Hint: $$\sum_{j=i}^{M}\frac{1}{\binom{j}{k}}=\frac{k}{k-1}\left(\frac{1}{\binom{i-1}{k-1}}-\frac{1}{\binom{M}{k-1}}\right)\Leftrightarrow\sum_{j=i}^{M}\frac{(j-k)!}{j!}=\frac{1}{k-1}\left(\frac{(i-k)!}{(i-1)!}-\frac{(M-k+1)!}{M!}\right)\quad \text{for } k\ge 2$$

c. Find the posterior mean.
d. Calculate the minimum number of observed tanks in order for the posterior mean to be finite.
e. Given $k=25$, $m=200$, and $\Omega=10000$, plot the posterior for the number n of tanks.

8. Further Study: When Will Civilization End? (10 Pts.)

Keywords: *Bayes Theorem, Doomsday Argument*

There are currently 7 billion people alive today, and according to the Population Reference Bureau, roughly 107 billion people have ever lived on planet earth. But is this data sufficient for us to estimate the total number of future humans, and more importantly, the time human species will go extinct, simply by using Bayes theorem? Well, some claim that it is.

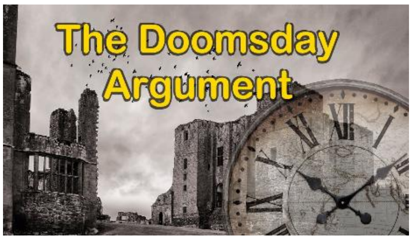


Figure 8 The Doomsday Argument asks a simple question: whether or not is it possible to apply probabilistic approaches to predict the total number of humans in the doomsday and the time it arrives?

This problem, known as **Doomsday Argument**, was initially raised by astrophysicist Brandon Carter in 1983, and since then various attempts have been made to address the issue. Despite different solutions proposed to date, no consensus has emerged on them so far.

In this problem, you are required to conduct a short survey and find a Bayesian solution for this problem. Your answer must include:

- An step-by-step solution as well as clear descriptions.
- The explicit definition of the terms and methods used in the process.
- The estimated population of total humans ever lived when humanity goes extinct.
- The time when humans go extinct in the future.
- Reference(s) used.

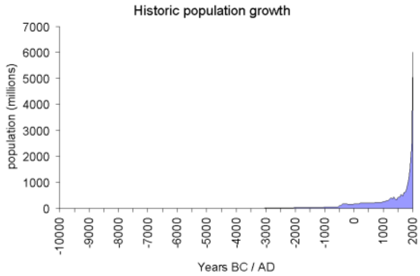


Figure 9 The growth of world population from 10,000 BC to AD 2000. A substantial increase over the past two centuries is noticeable.

Note 1: You are free to utilize relations and images from across the internet, but a mere translation wouldn't be of much worth.

Note 2: Your answers may not be totally accurate, but your efforts are worthwhile.

9. Some Explanatory Questions (8 Pts.)

Please answer the following questions as clear as possible:

a. In a binary classification problem, under what circumstances a Bayes classifier and a minimum distance classifier obtain exactly the same results?
b. Does a Minimum Distance Classifier have a training phase? What about a minimum-error classifier? Explain.
c. Assume a two-class 1D classification problem with the Gaussian distributions $p(\mathrm{x} \mid \omega_1) \sim N(-1,1)$ and $p(\mathrm{x} \mid \omega_2) \sim N(4,1)$, where the probabilities are equal. You are free to choose any classification method you would like, and you are given an infinitely large dataset. What would be the best error you can achieve on the test set, and why?
d. Discuss whether it is possible to plot ROC curve for a classification problem with more than two categories? If yes, how? And if no, why?
e. Is it possible to apply the Bayesian Decision Rule in a regression problem? If yes, explain how. If no, explain why.
f. Is the result of the Bayes decision rule unique? Explain.
g. When does MLE estimation lead to a better result than MAP estimation?
h. What is penalized MLE? When is it better to apply penalized MLE instead of normal MLE?

Good Luck!
Ali Abbasi