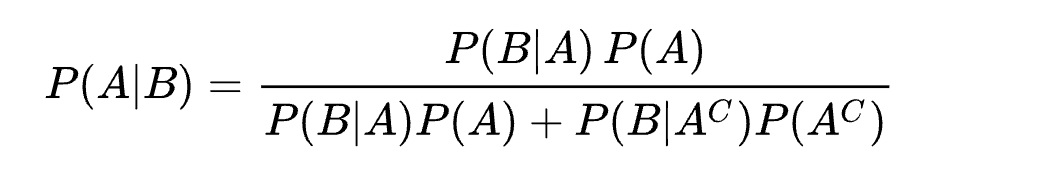
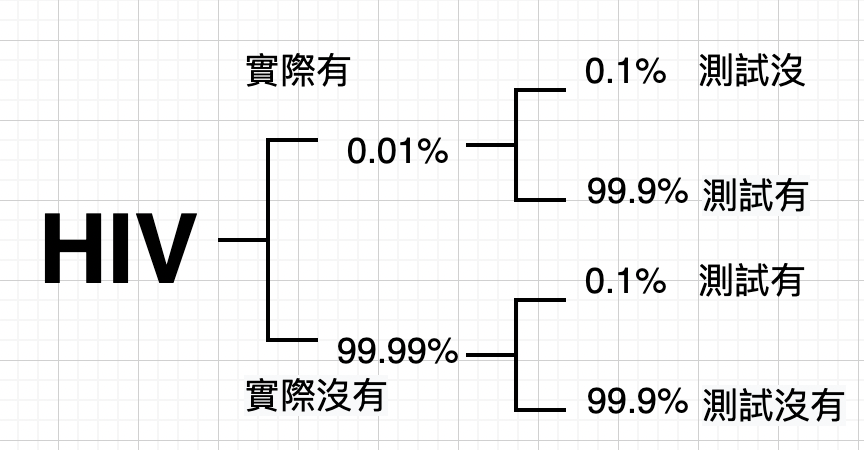
HW #2 Due: 3/28/2022

資工碩一 呂昀星 110598056

1. We mentioned the AIDS detection problem in the Bayesian decision theory. Use the Bayes theorem to confirm the given answer (i.e., 9%). To answer this problem, you need to distinguish two different conditions:
   * False positive is a conditional probability *P*(reagent is negative | patient is infected). Same argument for false negative.
   * When a patient is given a positive test result, it is actually *P*(patient is infected | reagent is positive)



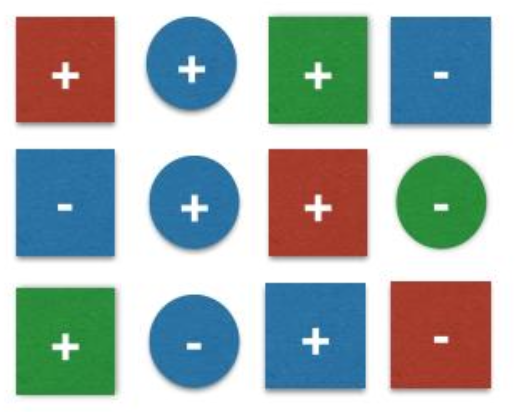


P(A)=實際有, P(B)=檢查有

1. We mentioned an example to use Naïve Bayesian classifier for classifying colored squares and circles in the lecture. Following the example, which class will

be assigned to?

Test pattern: blue circle



P(x|+)=P(B|+)+P(c|+)= 3/7 \* 2/7=6/49

P(x)=P(B)\*P(c)=6/12 \* 4/12= 1/6

P(+|x)=P(+)\*P(x|+)/P(x)=(7/12 \* 6/49)/(1/6)=0.4285

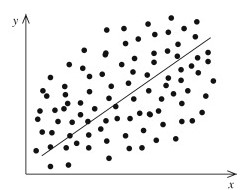
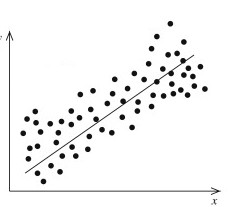
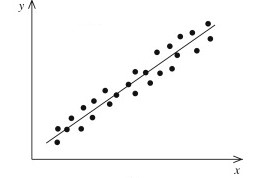
P(x|-)=P(B|-)+P(c|-)= 3/5 \* 2/5=6/25

P(x)=P(B)\*P(c)=6/12 \* 4/12= 1/6

P(-|x)=P(-)\*P(x|-)/P(x)=(5/12 \* 6/25)/(1/6)=**0.6**

Ans: it would be -

1. Below are scatterplots of Gaussian random points. Among these three plots, which one has the smallest correlation coefficient, and which one has the largest correlation coefficient? Is there any plot corresponding to negative correlation coefficient? Explain your answers.



(

a

)

(

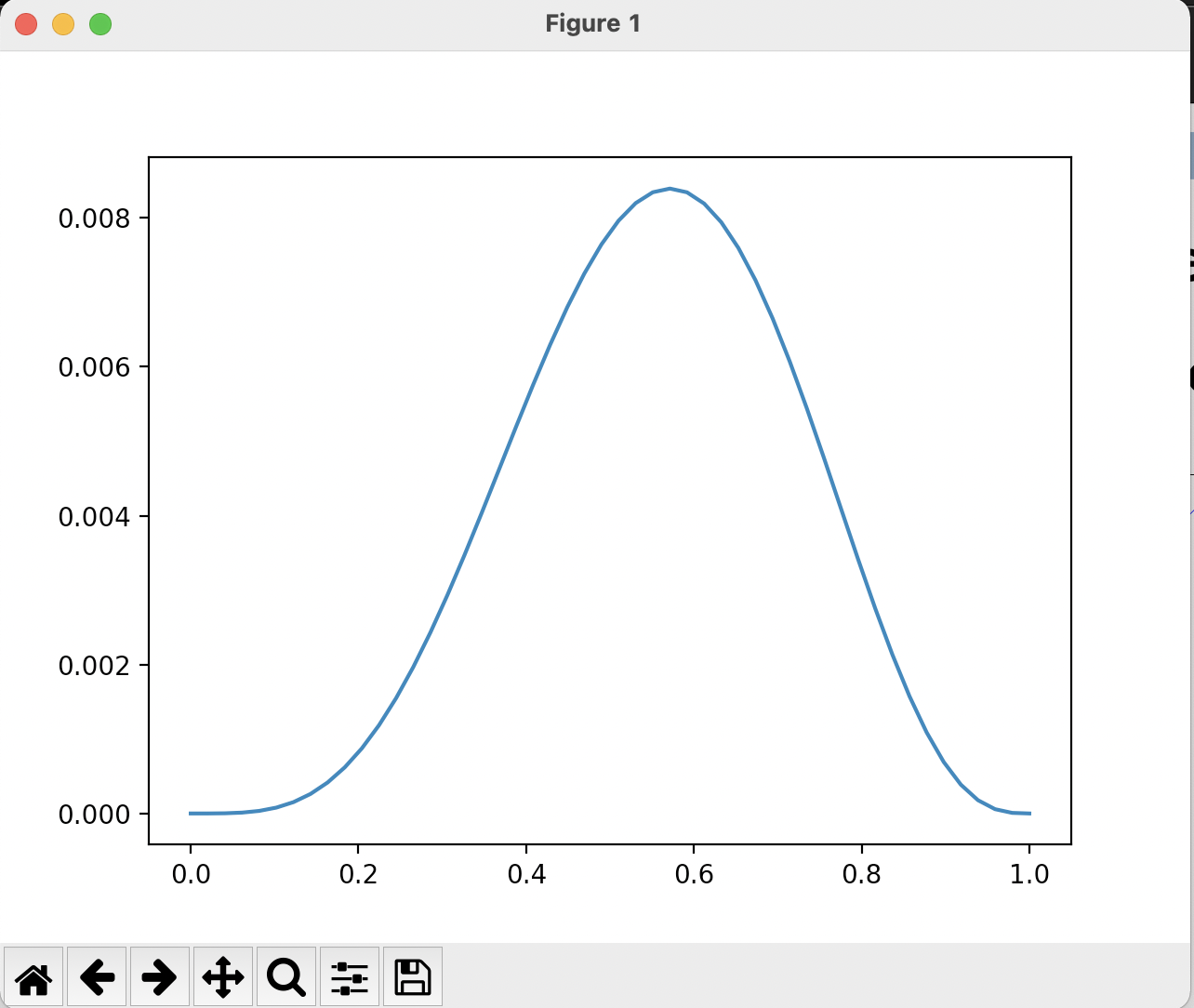
b) (c

)

Smallest: (a), Largest (c)

沒有負相關, 三個圖都正相關, 隨著x軸增加,y軸也增加

1. We mentioned the MLE in the lecture. Follow the lecture notes to write a program to draw the curve of probability with respect to 𝜃 = 𝑃({H}) for the coin sequence of {H,T,T,T,H,H,H}. Based on the curve, what is 𝜃̂, the ML estimate of 𝜃.
2. import matplotlib.pyplot as plt
3. import numpy as np
4. range = np.array([0,1])
5. def sequence(pro):
6. return pro\*(1-pro)\*(1-pro)\*(1-pro)\*pro\*pro\*pro
7. #{H,T,T,T,H,H,H}.
8. xnew = np.linspace(range.min(), range.max())
9. power\_smooth = sequence(xnew)
10. plt.plot(xnew, power\_smooth)
11. plt.show()



1. Use the Naïve Bayesian classifier (Note: Use GaussianNB because the features are continuous numbers) to classify the Iris dataset and compare the relative accuracy between this approach and the *k*-NN approach (in problem 4, HW #1).

Knn=0.97

Naïve Bayesian classifier=0.956

(knn is better)

from sklearn import datasets

from sklearn.model\_selection import train\_test\_split

iris = datasets.load\_iris()

#print(iris.feature\_names)

#print(iris.target\_names)

data\_x=iris.data

data\_y=iris.target

x\_train,x\_test,y\_train,y\_test=train\_test\_split(data\_x,data\_y,test\_size=0.3)

from sklearn.preprocessing import StandardScaler

sc = StandardScaler()

x\_train = sc.fit\_transform(x\_train)

x\_test = sc.transform(x\_test)

from sklearn.naive\_bayes import GaussianNB

classifier = GaussianNB()

classifier.fit(x\_train, y\_train)

y\_pred = classifier.predict(x\_test)

from sklearn.metrics import accuracy\_score

print ("Accuracy : ", accuracy\_score(y\_test, y\_pred))