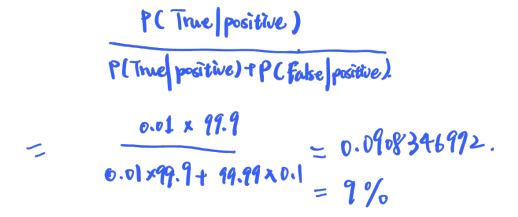
HW 02 Machine learning 111598067 王上澤

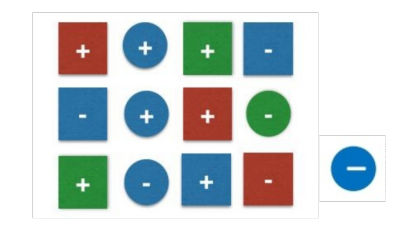
1. 0.1 % of false positive

0.1 % of false negative

0.01 % of people are the real HIV patient.







C∈{+, -}, X∈{X1, X2} where X1∈{R, G, B} X2∈{square, circle}

P(C=+) = 7/13, P(C=-) = 6/13, p(X1=R)=3/13, P(X1=G) = 3/13, P(X1=B) = 7/13, P(X2=S) = 8/13, P(X2=C) = 5/13, P(B | +) = 3/7,

For test pattern 

P(X|+) = P(B | +)P(S | +) = 3/7 \* 5/7 = 15/49

P(X|-) = P(B | -)P(S | -) = 4/6 \* 3/6 = 1/3

P(X) = P(B)P(S) = 7/13 \* 8/13 = 56/169

Therefore, P( + | X ) = P(+)P(X | +) / P(X) = 195/392

P( - | X ) = P(-)P(X | -) / P(X) = 13/28

Q1. Which one has the smallest correlation coefficient?

I think the answer is (C) , due to most of the points (data) is far from the line that tell the direction of data growth.

Q2. Which one has the largest correlation coefficient?

I think the answer is (a), due to most of the points is close to the line.

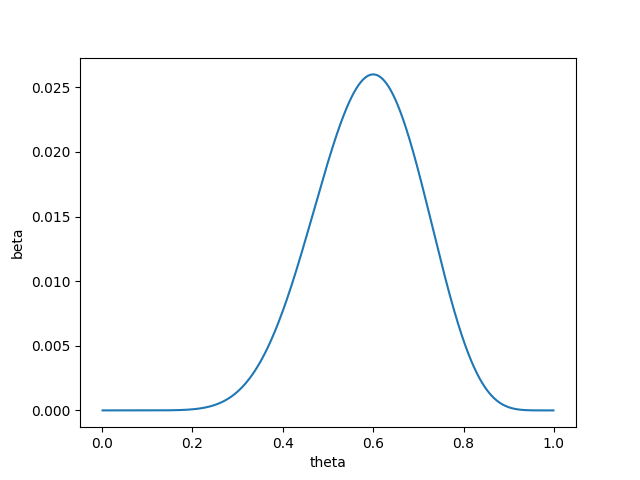
Q3. Is there any plot corresponding to negative correlation coefficient?

There is not negative correlation coefficient in three graphs, however , if there exist a plot that has negative correlation coefficient , it should look like the plot below.

一張含有 圖表 的圖片

自動產生的描述

1. max\_theta = 0.6 max\_beta = 0.026005292974080003



一張含有 文字 的圖片

自動產生的描述

一張含有 文字 的圖片

自動產生的描述

一張含有 文字 的圖片

自動產生的描述

1. the average accuracy of 10 times trial = 0.9444444444444444

I remove 'mean compactness', 'compactness error', 'worst compactness' columns in dataset due to this attribute is calculated by another attributes. And the result shows that the accuracy still high after I remove the columns.

一張含有 文字 的圖片

自動產生的描述

I delete the rows that missing the attribute. (by pandas dataframe dropna() method)