CS464 Introduction to Machine Learning
Homework Assignment 1

Ilker Demirel

21502856

Q1

Q1.1

Probability of any two students having same birthday in a classroom of size n is (call this event, event A),

$$\begin{aligned} P(A) &= 1 - P(A') \\ &= 1 - \frac{365 \cdot 364 \cdot \cdots \cdot (365 - (n-1))}{365^n} \end{aligned} \tag{1}$$

In equation (1), the denominator of the subtracted term is the total number of birthday combinations, and the nominator is the total number of birthday combinations when everyone student has a different birthday. See figure (1) for n vs $P(A)$.

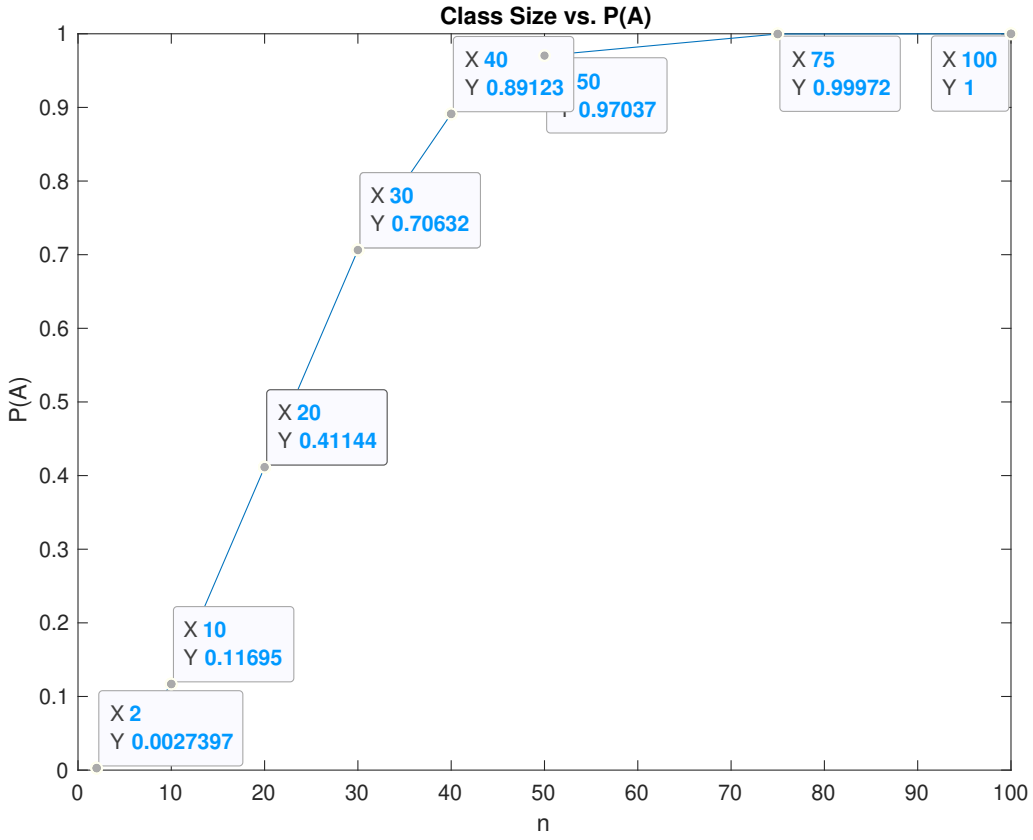

Figure 1: Class Size (n) vs. $P(A)$

Q1.2

Minimum number of students to make sure that any two students have the same birthday is 366, since there are 365 days in a year.

Q2

Q2.1

$P(S = disease) = 0.011$
$P(S = healthy) = 0.989$
$P(T = positive|S = disease) = 0.94$
$P(T = negative|S = disease) = 0.06$
$P(T = positive|S = healthy) = 0.02$
$P(T = negative|S = healthy) = 0.98$

Q2.2

$P(S = disease|T = positive) = \frac{P(T=positive|S=disease)P(S=disease)}{P(T=positive)} = \frac{0.94 \cdot 0.011}{0.94 \cdot 0.011 + 0.02 \cdot 0.989} \approx 0.35$

Given the test result for a patient is positive, the probability that patient has the disease is 0.35. Therefore it is not reasonable to diagnose a patient with the disease when the test result is positive.

Q2.3

Since $P(S = disease|T = positive) \neq 1$, one can **never** definitely diagnose a patient as sick. That being said, the criterion for **confidentally** diagnosing a patient as sick is not defined in the question. Given there are n positive test results, the probability that the patient is sick is,

$$P(S = disease|n\ positive\ tests) = 1 - 0.65^n \tag{2}$$

According to equation (2), if there are seven positive tests, the patient is sick with a probability of $1 - 0.65^7 \approx 0.95$.

Q3