

## 1a: Manually calculate prediction using the Naive Bayes Model and K nearest neighbor for the test example you are assigned

The below image is the Naive Bayes Model for the "A gentle reminder about the team's presentation tomorrow" --> [0,0,1,0,1]

$$\begin{aligned}
 \underline{A} &= P(\text{Spam} | E) = P(E | \text{Spam}) \cdot P(\text{Spam}) \\
 &= P(\text{win} | \text{Spam}) \cdot P(\text{click} | \text{Spam}) \cdot P(\text{team} | \text{Spam}) \cdot P(\text{claim} | \text{Spam}) \cdot P(\text{reminder} | \text{Spam}) \cdot P(\text{Spam}) \\
 &= \frac{3}{7} \times \frac{6}{7} \times \frac{1}{7} \times \frac{4}{7} \times \frac{1}{9} \times \frac{7}{15} = 0.0015 \\
 &\quad \uparrow \\
 &\quad \text{Applied Smoothing}
 \end{aligned}$$

$$\begin{aligned}
 \underline{B} &= P(\text{Ham} | E) = P(E | \text{Ham}) \cdot P(\text{Ham}) \\
 &= P(\text{win} | \text{Ham}) \cdot P(\text{click} | \text{Ham}) \cdot P(\text{team} | \text{Ham}) \cdot P(\text{claim} | \text{Ham}) \cdot P(\text{reminder} | \text{Ham}) \cdot P(\text{Ham}) \\
 &= \frac{8}{8} \times \frac{4}{8} \times \frac{6}{8} \times \frac{8}{8} \times \frac{3}{8} \times \frac{8}{15} = 0.075
 \end{aligned}$$

$\therefore P(\text{Ham} | E)$  is greater than  $P(\text{Spam} | E)$  so classification is "Ham" for the above test data.

$P(\text{Ham} | E) \cdot P(\text{Ham}) = 0.075$  is greater than the  $P(\text{Spam} | E) \cdot P(\text{Spam}) = 0.0015$

So the message A gentle reminder about the team's presentation tomorrow is classified as "Ham"

## 1(b) The below images are the K nearest neighbor for A gentle reminder about the team's presentation tomorrow where K=3.

Calculating the Euclidean distance for the above test data.

Test data = {0, 0, 1, 0, 1}

- ① Train = [1, 1, 0, 0, 0]  

$$d = \sqrt{(0-1)^2 + (0-1)^2 + (1-0)^2 + (0-0)^2 + (1-0)^2}$$

$$= \sqrt{1+1+1+0+1} = \sqrt{4} = 2$$
- ② Train = [0, 1, 1, 0, 0]  

$$d = \sqrt{(0-0)^2 + (0-1)^2 + (1-1)^2 + (0-0)^2 + (1-0)^2}$$

$$= \sqrt{0+1+0+0+1} = \sqrt{2} = 1.41$$
- ③ Train = [0, 0, 0, 1, 0]  

$$d = \sqrt{(0-0)^2 + (0-0)^2 + (1-0)^2 + (0-1)^2 + (1-0)^2}$$

$$= \sqrt{0+0+1+1+1} = \sqrt{3} = 1.73$$
- ④ Train = [0, 0, 1, 0, 1]  

$$d = \sqrt{(0-0)^2 + (0-0)^2 + (1-1)^2 + (0-0)^2 + (1-1)^2}$$

$$= \sqrt{0+0+0+0+0} = 0$$

$$\textcircled{5} \text{ Train} = [1, 0, 0, 0, 0]$$

$$d = \sqrt{(0-1)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2 + (1-0)^2}$$

$$= \sqrt{1+0+1+0+1} = \sqrt{3} = 1.73$$

$$\textcircled{6} \text{ Train} = [0, 0, 0, 0, 1]$$

$$d = \sqrt{(0-0)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2 + (1-1)^2}$$

$$= \sqrt{0+0+1+0+0} = \sqrt{1} = 1$$

$$\textcircled{7} \text{ Train} = [1, 0, 1, 0, 0]$$

$$d = \sqrt{(0-1)^2 + (0-0)^2 + (1-1)^2 + (0-0)^2 + (1-0)^2}$$

$$= \sqrt{1+0+0+0+1} = \sqrt{2} = 1.41$$

$$\textcircled{8} \text{ Train} = [0, 1, 1, 0, 0]$$

$$d = \sqrt{(0-0)^2 + (0-1)^2 + (1-1)^2 + (0-0)^2 + (1-0)^2}$$

$$= \sqrt{0+1+0+0+1} = \sqrt{2} = 1.41$$

$$\textcircled{9} \text{ Train} = \cancel{[0, 0, 0, 0, 1]} [0, 0, 0, 1, 0]$$

$$d = \sqrt{(0-0)^2 + (0-0)^2 + (1-0)^2 + (0-1)^2 + (1-0)^2}$$

$$= \sqrt{0+0+1+1+1} = \sqrt{3} = 1.73$$

$$\textcircled{10} \text{ Train} = [0, 1, 1, 0, 0]$$

$$d = \sqrt{(0-0)^2 + (0-1)^2 + (1-1)^2 + (0-0)^2 + (1-0)^2}$$

$$= \sqrt{0+1+0+0+1} = \sqrt{2} = 1.41$$

$$\textcircled{11} \text{ Train} = [0, 0, 1, 0, 0]$$

$$d = \sqrt{(0-0)^2 + (0-0)^2 + (1-1)^2 + (0-0)^2 + (1-0)^2}$$

$$= \sqrt{0+0+0+0+1} = 1$$

$$\textcircled{12} \text{ Train} = [1, 0, 0, 0, 0]$$

$$d = \sqrt{(0-1)^2 + (0-0)^2 + (1-0)^2 + (0-0)^2 + (0-0)^2}$$

$$= \sqrt{1+0+1+0+0} = \sqrt{2} = 1.41$$

$$\textcircled{13} \text{ Train} = [0, 1, 0, 0, 0]$$

$$d = \sqrt{(0-0)^2 + (1-0)^2 + (1-0)^2 + (0-0)^2 + (0-1)^2}$$

$$= \sqrt{0+1+1+0+1} = \sqrt{3} = 1.73$$

$$\textcircled{14} \text{ Train} = [0, 0, 0, 1, 0]$$

$$d = \sqrt{(0-0)^2 + (0-0)^2 + (1-0)^2 + (0-1)^2 + (1-0)^2}$$

$$= \sqrt{0+0+1+1+1} = \sqrt{3} = 1.73$$

$$\textcircled{15} \text{ Train} = [0, 0, 1, 0, 1]$$

$$d = \sqrt{(0-0)^2 + (0-0)^2 + (1-1)^2 + (0-0)^2 + (1-1)^2}$$

$$= \sqrt{0+0+0+0+0} = 0$$

The 3NN for "A gentle reminder about the team's presentation tomorrow" are listed below:

"A reminder for tomorrow's team-building exercise."

"The reminder for project deadlines is attached."

"Reminder: Year-end team gathering is on the 20<sup>th</sup>"

All three nearest neighbors are classified as "Ham". Therefore, the predicted classification for the test message "A gentle reminder about the team's presentation tomorrow" is also "Ham".