1. x_8 left

Hamming distance:

$d(x_1, x_8) = d(A0, B1) = 2$ $K = 5:$

$d(x_2, x_8) = d(B1, B1) = 0$ x_2, x_3, x_5, x_6, x_7

$d(x_3, x_8) = d(A1, B1) = 1$ P P N N N

$d(x_4, x_8) = d(A0, B1) = 2$ x_8 is classified

$d(x_5, x_8) = d(B0, B1) = 1$ as negative

$d(x_6, x_8) = d(B0, B1) = 1$ (TN)

$d(x_7, x_8) = d(A1, B1) = 1$

x_7 left

$d(x_1, x_7) = d(A0, A1) = 1$ $d(x_5, x_7) = d(B0, A1) = 2$

$d(x_2, x_7) = d(B1, A1) = 1$ $d(x_6, x_7) = d(B0, A1) = 2$

$d(x_3, x_7) = d(A1, A1) = 0$ $d(x_8, x_7) = d(B1, A1) = 1$

$d(x_4, x_7) = d(A0, A1) = 1$ x_1, x_2, x_3, x_4, x_8

                               P, P, P, P, N

x_7 is classified as positive   (FP)

x_6 left

$d(x_1, x_6) = 1$ $d(x_5, x_6) = 0$ x_1, x_2, x_4, x_5, x_8

$d(x_2, x_6) = 1$ $d(x_4, x_6) = 2$ P P P N N

$d(x_3, x_6) = 2$ $d(x_8, x_6) = 1$

$d(x_4, x_6) = 1$ x_6 is classified as positive

x_5 left   (FP)

$d(x_1, x_5) = 1$ $d(x_2, x_5) = 1$ $d(x_3, x_5) = 2$

$d(x_4, x_5) = 1$ $d(x_6, x_5) = 0$ $d(x_7, x_5) = 2$

$d(x_8, x_5) = 1$ x_1, x_2, x_4, x_6, x_8

                      P P P N N

         x_5 is classified as positive

x_4 left   (FP)

$d(x_1, x_4) = 0$ $d(x_2, x_4) = 2$ $d(x_3, x_4)$ 1 $d(x_5, x_4) = 1$

$d(x_6, x_4) = 1$ $d(x_7, x_4) = 1$ $d(x_8, x_4) = 2$

x_1, x_3, x_5, x_6, x_7 x_4 is classified as

 P   P   N   N   N   negative (FN)

x_3 left

$d(x_1, x_3) = 1$ $d(x_2, x_3) = 1$ $d(x_4, x_3) = 1$

$d(x_5, x_3) = 2$ $d(x_6, x_3) = 2$ $d(x_7, x_3) = 0$

$d(x_8, x_3) = 1$ x_1, x_2, x_4, x_7, x_8

                      P P P N N

   x_3 is classified as positive   (TP)

x_2 left

$d(x_1, x_2) = 2$ $d(x_3, x_2) = 1$ $d(x_4, x_2) = 2$

$d(x_5, x_2) = 1$ $d(x_6, x_2) = 1$ $d(x_7, x_2) = 1$ $d(x_8, x_2) = 0$

x_2, x_5, x_6, x_7, x_8 x_2 is classified

 P   N   N   N   N   as negative

                              (FN)

x_1 left

$d(x_2, x_1) = 2$ $d(x_3, x_1) = 1$ $d(x_4, x_1) = 0$

$d(x_5, x_1) = 1$ $d(x_6, x_1) = 1$ $d(x_7, x_1) = 1$ $d(x_8, x_1) = 2$

x_3, x_4, x_5, x_6, x_7 x_1 is classified

 P   P   N   N   N   as negative

                              (FN)

TP = 1

TN = 1

FP = 3   Precision = $\dfrac{TP}{TP + FP}$ = $\dfrac{1}{1+3}$ = 1/4

FN = 3

Recall = $\dfrac{TP}{TP + FN}$ = $\dfrac{1}{1+3}$ = 1/4

$F = 2 \cdot \dfrac{Precision \cdot Recall}{Precision + Recall}$ = $2 \cdot \dfrac{1/16}{1/2}$ =

= $2 \cdot \dfrac{2}{16}$ = $\dfrac{4}{16}$ = $\dfrac{1}{4}$

$\dfrac{\frac{1}{2}}{\frac{1}{5}}$ = $\dfrac{5}{2}$ = 2.5

2. K=5, that means the model necessarily sees at least 2 observation with different outcomes as $similar$. In all cases we have 1 neighbour as identical and 4 with distance 1. To prevent ties, we will try $k=1$.

x_8 left x_2 P (N)
x_4 left x_3 P (N)
x_6 left x_5 N (N)
x_5 left x_6 N (N)
x_4 left x_1 P (P)
x_3 left x_4 N (P)
x_2 left x_8 N (P)
x_1 left x_4 P (P)

$TP = 2$ $TN = 2$ $FP = 2$ $FN = 2$

$$Precision = \frac{2}{2+2} = \frac{1}{2} \quad Recall = \frac{2}{2+2} = \frac{1}{2}$$

$$F_1 = 2 \cdot \frac{P \cdot R}{P+R} = 2 \cdot \frac{(\frac{1}{2})^2}{\frac{1}{2}+\frac{1}{2}} =$$

$$= \frac{\frac{1}{2}}{1} = \frac{1}{2} \quad increase \ of \ 2 \ fold$$

We can see in the dataset that y_1 is a better predictor than y_2, what if we apply hamming distance where only y_1 counts and $k=3$

x_8 left
$d(x_1,x_8) = 1$ $d(x_2,x_8)=0$ $d(x_3,x_8)=1$
$d(x_4,x_8) = 1$ $d(x_5,x_8) = 0$ $d(x_6,x_8)=0$
$d(x_7,x_8) = 1$ (TN) $TN = 2$
                          $FP = 1$

x_7 left
$d(x_1,x_7) = 0$ $d(x_2,x_7) = 1$ $d(x_3,x_7)=0$
$d(x_4,x_7)=0$ $d(x_5,x_7)=1$ $d(x_6,x_7)=1$
$d(x_8,x_7)= 1$ x_1, x_3, x_4 (FP)
                 P   P   P

$TN=2$
$FP=4$

x_6 left $d(x_1,x_6)=1$ $d(x_2,x_6)=0$
$d(x_3,x_6)=1$ $d(x_4,x_6)=1$ $d(x_5,x_6)=0$
$d(x_7,x_6)=1$ $d(x_8,x_6)=0$ $x_2,(x_7,x_8)$
                           P   N   N
                            (TN)

x_5 left
$d(x_1,x_5)=1$ $d(x_2,x_5) = 0$ $d(x_3,x_5) = 1$
$d(x_4,x_5) = 1$ $d(x_6,x_5)=0$ $d(x_7,x_5)=1$
$d(x_8,x_5)=0$ x_2,x_6,x_6 are neighbours
                 P   N   N   (TN)

x_4 left
$d(x_1,x_4) = 0$ $d(x_2,x_4) = 1$ $d(x_3,x_4)=0$
$d(x_5,x_4) = 1$ $d(x_6,x_4) = 1$ $d(x_7,x_4)=0$
$d(x_8,x_4)=1$ x_1,x_3,x_7 (TP)
                 P   P   N

x_3 left
$d(x_1,x_3) = 0$ $d(x_2,x_3) = 1$ $d(x_4,x_3)=0$
$d(x_5,x_3) = 1$ $d(x_6,x_3) = 1$ $d(x_7,x_3)=0$
$d(x_8,x_3)=1$ x_1,x_4,x_7 (TP)
                 P   P   N

x_2 left
$d(x_1,x_2) = 1$ $d(x_3,x_2)=1$ $d(x_4,x_2)=1$
$d(x_5,x_2) = 0$ $d(x_6,x_2) = 0$ $d(x_7,x_2)=1$
$d(x_8,x_2)=0$ x_5,x_6,x_8 (FN)
                 N   N   N

x_1 left
$d(x_2,x_1) = 1$ $d(x_3,x_1)=0$ $d(x_4,x_1)=0$
$d(x_5,x_1) = 1$ $d(x_6,x_1) = 1$ $d(x_7,x_1)=0$ $d(x_8,x_1)=1$
x_3,x_4,x_7 (TP)
P   P   N

$TP=3$ $TN=3$ $FP = 1$ $FN=1$

$$Precision = \frac{3}{3+1} = 3/4$$

$$Recall = \frac{3}{3+1} = 3/4$$

$$F_1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} = 2 \cdot \frac{(3/4)^2}{\frac{3+3}{4}}$$

$$= 2 \cdot \frac{\frac{9}{16}}{\frac{6}{4}} = 2 \cdot \frac{36}{16 \cdot 6} = \frac{72}{32 \cdot 3}$$

$$= \frac{72}{96} = \frac{3}{4} \quad Increase \ of \ 3 \ fold$$

3. $P(c=p) = \frac{5}{9}$ $P(c=N) = \frac{4}{9}$

$$P(c=p \mid know) = \frac{P(y_1, y_2, y_3 \mid c) \cdot P(c)}{P(y_1, y_2, y_3)} =$$

$$= \frac{P(y_1, y_2 \mid c) \cdot P(y_3 \mid c) \cdot P(c)}{P(y_1, y_2) \cdot P(y_3)} =$$

$$= \frac{P(y_1, y_2 \mid c) \cdot P(c)}{P(y_1, y_2)} \cdot \frac{P(y_3 \mid c)}{P(y_3)} = \quad \begin{array}{l} \text{todos dividem por} \\ P(y_1, y_2, y_3) \end{array}$$

$\underbrace{}_{2^2 \text{ combinações}}$ $\qquad \underbrace{}_{\text{normal}}$

$$= P(y_1, y_2 \mid c) \cdot P(c) \cdot P(y_3 \mid c) =$$

$$= P(y_1, y_2 \mid c) \cdot P(c) \cdot \left(\frac{1}{\sqrt{2\pi}\sigma} \exp\left\{ -\frac{1}{2\sigma^2}(x-\mu)^2 \right\} \right)$$

$P(y_1 = A, y_2 = 0 \mid c=p) = \frac{2}{5}$ $\qquad$ $P(y_1 = A, y_2 = 0 \mid c=N) = 0$

$P(y_1 = B, y_2 = 0 \mid c=p) = \frac{1}{5}$ $\qquad$ $P(y_1 = B, y_2 = 0 \mid c=N) = \frac{1}{2}$

$P(y_1 = A, y_2 = 1 \mid c=p) = 1/5$ $\qquad$ $P(y_1 = A, y_2 = 1 \mid c=N) = \frac{1}{4}$

$P(y_1 = B, y_2 = 1 \mid c=p) = \frac{1}{5}$ $\qquad$ $P(y_1 = B, y_2 = 1 \mid c=N) = \frac{1}{4}$

Para $c = p$; $\mu_{y_3} = 0.82$

$$\sigma_{y_3} = \sqrt{\frac{1}{4} \sum_1^5 (x_i - 0.82)^2} = 0.217$$

$$\sigma_{y_3}^2 = 0.047$$

Para $c = N$; $\mu_{y_3} = 1$

$$\sigma_{y_3} = \sqrt{\frac{1}{3} \sum_1^4 (x_i - 1)^2} = 0.122$$

$$\sigma_{y_3}^2 = 0.01488$$

Classifier:

A0:

$C = P$: $\frac{2}{5} \cdot \frac{5}{9} \left(\underbrace{\frac{1}{0.217\sqrt{2\pi}}}_{1.8384} \exp\left(\frac{1}{2 \cdot 0.047}(x-0.82)^2 \right) \right)$

$C = N$: 0 $\qquad\qquad\qquad\qquad$ $\frac{1}{0.122\sqrt{2\pi}} = 3.27$

A1:

$C = P$: $\frac{1}{5} \cdot \frac{5}{9} \left(1.8384 \cdot \exp\left\{ -10.638(x-0.82)^2 \right\} \right)$

$C = N$: $\frac{1}{4} \cdot \frac{4}{9} \left(3.27 \exp\left\{ -33.602(x-1)^2 \right\} \right)$

B0:

$C = P$: $\frac{1}{5} \cdot \frac{5}{9} \left(1.8384 \exp\left\{ -10.638(x-0.82)^2 \right\} \right)$

$C = N$: $\frac{1}{2} \cdot \frac{4}{9} \left(3.27 \exp\left\{ -33.602(x-1)^2 \right\} \right)$

B1:

$C = P$: $\frac{1}{5} \cdot \frac{5}{9} \left(1.8384 \exp\left\{ -10.638(x-0.82)^2 \right\} \right)$

$C = N$: $\frac{1}{4} \cdot \frac{4}{9} \left(3.27 \exp\left\{ -33.02(x-1)^2 \right\} \right)$

4. $\{(A;1;0.8),(B;1;1),(B;0;0.9)\}$

MAP → max a posteriori.

$\alpha(A1):$

$C=P:\ \frac{1}{5}\cdot\frac{5}{9}\left(1.8384\,exp\left\{-10.638\,(0.8-0.82)^2\right\}\right)=$

$\qquad =\boxed{0.1108}$

$C=N:\ \frac{1}{4}\cdot\frac{4}{9}\left(3.27\,exp\left\{-33.802\,(0.8-1)^2\right\}\right)=$

$\qquad =0.0226$

$\beta(B1).$

$C=P:\ \frac{1}{5}\cdot\frac{5}{9}\left(1.8384\,exp\left\{-10.638\,(1-0.82)^2\right\}\right)=$

$\qquad =\ 0.09008$

$C=N:\ \frac{1}{4}\cdot\frac{4}{9}\left(3.27\,exp\left\{-33.602\,(1-1)^2\right\}\right)$

$\qquad =\ \boxed{0.41}$

$\gamma(B0):$

$C=P:\ \frac{1}{5}\cdot\frac{5}{9}\left(1.8384\,exp\left\{-10.638\,(0.9-0.82)^2\right\}\right)$

$\qquad =0.1066$

$C=N:\ \frac{1}{2}\cdot\frac{4}{9}\left(3.27\,exp\left\{-33.602\,(0.9-1)^2\right\}\right)=$

$\qquad =\boxed{0.1492}$

$\alpha \to P \qquad \beta \to N \qquad \gamma \to N$

5. Naive Bayes: $P(y_1, y_2, \dots, y_n | c) =$
$= P(y_1|c) \cdot P(y_2|c) \cdot \dots \cdot P(y_n|c)$

TIL assumption: $p(u_1) = p(u_2) = \dots = p(u_n)$ ∴

classifier: $h_{ML} = \text{argmax } P(D|h) \, P(h) \Rightarrow$

$\Rightarrow h_{ML} = \text{argmax } P(D|h)$

The goal is to classify the sentence
$$\text{"I like to Run"}$$
$$t_1 \quad t_2 \quad t_3 \quad t_4$$

$\{$ ("Amazing Run", P), ("I like it", P),
("Too tired", N), ("Bad run", N) $\}$

$$P(c|x_{new}) = \frac{P(t_1, t_2, t_3, t_4 | c) \cdot P(c)}{P(t_1, t_2, t_3, t_4)} =$$

$$= p(t_1|c) \cdot p(t_2|c) \cdot p(t_3|c) \cdot p(t_4|c) \cdot P(c)$$

$$= p(t_1|c) \cdot p(t_2|c) \cdot p(t_3|c) \cdot p(t_4|c)$$

para $\boxed{c = P}$: $p(t_1|c=p) \cdot p(t_2|c=p) \cdot p(t_3|c=p) \cdot p(t_4|c=p)$

$\text{freq}(t_1) = 1$ $\qquad\qquad = \frac{1+1}{5+4} \cdot \frac{1+1}{5+4} \cdot \frac{0+1}{3+4} \cdot \frac{1+1}{5+4} =$
$\text{freq}(t_2) = 1 \qquad v = 4$
$\text{freq}(t_3) = 0 \qquad N_c = 5 \qquad = 2.082 \cdot 10^{-4}$
$\text{freq}(t_4) = 1$

$\qquad\qquad\qquad\qquad\qquad\qquad \text{freq}(t_1) = 0$
$\qquad\qquad\qquad\qquad\qquad\qquad \text{freq}(t_2) = 0 \qquad v = 4$
para $\boxed{c = N}$: $\qquad\qquad \text{freq}(t_3) = 0 \qquad N_c = 4$
$\qquad\qquad\qquad\qquad\qquad\qquad \text{freq}(t_4) = 1$

$p(t_1|c=N) \cdot p(t_2|c=N) \cdot p(t_3|c=N) \cdot p(t_4|c=N) =$

$$= \frac{0+1}{4+4} \cdot \frac{0+1}{4+4} \cdot \frac{0+1}{4+4} \cdot \frac{1+1}{4+4} = 7 \cdot 10^{-5}$$

I like to run classificado como P