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

$d(x_2,x_6) = 1$ $d(x_7,x_6) = 2$ 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

FP

x_5 left

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

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

$d(x_8,x_5) = 1$ 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

FN = 3

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

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

$$F_1 = 2 \cdot \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} = 2 \cdot \frac{1/16}{1/2} =$$

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

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

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_7 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_7 N	(P)
x_2 left	x_8 N	(P)
x_1 left	x_4 P	(P)

$TP = 2 \quad TN = 2 \quad FP = 2 \quad FN = 2$

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

$$f_1 = 2 \cdot \frac{P \cdot R}{P + R} = 2 \cdot \frac{(\frac{1}{2})^2}{\frac{1}{2}+\frac{1}{2}} =$$

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

We can see in the data set 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

$d(x_1,x_8) = 1 \quad d(x_2,x_8) = 0 \quad d(x_3,x_8) = 1$

$d(x_4,x_8) = 1 \quad d(x_5,x_8) = 0 \quad d(x_6,x_8) = 0$

$d(x_7,x_8) = 1$

TN

$TN = 2$

$FP = 1$

x_7 left

$d(x_1,x_7) = 0 \quad d(x_2,x_7) = 1 \quad d(x_3,x_7) = 0$

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

$d(x_8,x_7) = 1$

x_1, x_3, x_4

P P P

FP

$TN = 2$

$FP = 4$

x_6 left $d(x_1,x_6) = 1 \quad d(x_2,x_6) = 0$

$d(x_3,x_6) = 1 \quad d(x_4,x_6) = 1 \quad d(x_5,x_6) = 0$

$d(x_7,x_6) = 1 \quad d(x_8,x_6) = 0$

x_2, x_5, x_8

P N N

TN

x_5 left

$d(x_1,x_5) = 1 \quad d(x_2,x_5) = 0 \quad d(x_3,x_5) = 1$

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

$d(x_8,x_5) = 0$

x_2, x_6, x_8 are neighbours

P N N

TN

x_4 left

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

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

$d(x_8,x_4) = 1$

x_1, x_3, x_7

P P N

TP

x_3 left

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

$d(x_5,x_3) = 1 \quad d(x_6,x_3) = 1 \quad d(x_7,x_3) = 0$

$d(x_8,x_3) = 1$

x_1, x_4, x_7

P P N

TP

x_2 left

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

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

$d(x_8,x_2) = 0$

x_5, x_6, x_8

N N N

FN

x_1 left

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

$d(x_5,x_1) = 1 \quad d(x_6,x_1) = 1 \quad d(x_7,x_1) = 0 \quad d(x_8,x_1) = 1$

x_3, x_4, x_7

P P N

TP

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

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

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

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

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

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

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

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

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

$$= \frac{P(y_1, y_2 \mid c) \cdot P(c)}{\underset{2^2 \text{ combinações}}{P(y_1, y_2)}} \cdot \frac{P(y_3 \mid c)}{\underset{\text{normal}}{P(y_3)}} =$$

todas dividem por $P(y_1, y_2, y_3)$

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

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

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

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

$P(y_1 = A, y_2 = 1 \mid c=P) = 1/6$ $P(y_1 = A, y_2 = 1 \mid c=N) = \frac{1}{4}$

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

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 $c=N$; $\mu_{y_3} = 0.82$

$$\sigma_{y_3} = \sqrt{\frac{1}{3} \sum_1^5 (x_i - 1)^2} = 0.122$$

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

Classificar:

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

$\frac{1}{0.122\sqrt{2\pi}} = 3.27$

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.

$\alpha(A1)$:

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

$= \boxed{0.1108}$

$c = N: \frac{1}{4} \cdot \frac{4}{9} (3.27 \exp\{-33.602 (0.8 - 1)^2\}) =$

$= 0.0226$

$\beta(B1)$.

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

$= 0.09008$

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

$= \boxed{0.(1)}$

$\gamma(B0)$:

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

$= 0.1066$

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

$= \boxed{0.1492}$

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

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

$= P(y_1|c) \cdot P(y_2|c) \cdot \ldots \cdot P(y_n|c)$

T/L assumption: $p(h_1) = p(h_2) = \ldots = p(h_n)$ ∴

classifier: $h_{ML} = \arg\max P(D|h) P(h) \Rightarrow$

$\Rightarrow h_{ML} = \arg\max P(D|h)$

The goal is to classify the sentence "I like to Run

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

$freq(t_1) = 1$

$freq(t_2) = 1$ $\quad V = 9$

$freq(t_3) = 0$ $\quad N_c = 5$

$freq(t_4) = 1$

$$= \frac{1+1}{5+9} \cdot \frac{1+1}{5+9} \cdot \frac{0+1}{3+9} \cdot \frac{1+1}{5+9} =$$

$$= 2.082 \cdot 10^{-4}$$

para $\boxed{c = N}$:

$freq(t_1) = 0$ $\quad V = 9$

$freq(t_2) = 0$ $\quad N_c = 4$

$freq(t_3) = 0$

$freq(t_4) = 1$

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

$$= \frac{0+1}{4+9} \cdot \frac{0+1}{4+9} \cdot \frac{0+1}{4+9} \cdot \frac{1+1}{4+9} = 7 \cdot 10^{-5}$$

I like to run classificado como P