I. Pen-and-paper

(1.) Hamming distance: $d(x_i, x_j)$ = número de features diferentes.

Temos então a tabela de distâncias $d(x_i, x_j)$:

$i \backslash j$	1	2	3	4	5	6	7	8
1	0	2	1	0	1	1	1	2
2	2	0	1	2	1	1	1	0
3	1	1	0	1	2	2	0	1
4	0	2	1	0	1	1	1	2
5	1	1	2	1	0	0	2	1
6	1	1	2	1	0	0	2	1
7	1	1	0	1	2	2	0	1
8	2	0	1	2	1	1	1	0

Para x_1:

A k-vizinhança é $\{x_3, x_4, x_5, x_6, x_7\}$.

$$f(x_1) = \underset{c \in Z}{\operatorname{argmax}} \sum_{i=1}^{k} \delta(c, f(x_i))$$

Para $c = P$:

$$\sum_{i=1}^{k} \delta(c, f(x_i)) = 2$$

Para $c = N$:

$$\sum_{i=1}^{k} \delta(c, f(x_i)) = 3$$

Logo, x_1 deve ser classificado como N.

Para x_2:

A k-vizinhança é $\{x_3, x_5, x_6, x_7, x_8\}$.

$$f(x_2) = \underset{c \in Z}{\operatorname{argmax}} \sum_{i=1}^{k} \delta(c, f(x_i))$$

Para $c = P$:

$$\sum_{i=1}^{k} \delta(c, f(x_i)) = 1$$

para $c = N$:

$$\sum_{i=1}^{k} \delta(c, f(x_i)) = 4$$

Logo, x_2 deverá ser classificado como N.

Para x_3:

A k-vizinhança é $\{x_1, x_2, x_4, x_7, x_8\}$.

$$f(x_3) = \underset{c \in Z}{\operatorname{argmax}} \sum_{i=1}^{k} \delta(c, f(x_i))$$

Para $c = P$:

$$\sum_{i=1}^{K} \delta(c, f(x_i)) = 3$$

Para $c = N$:

$$\sum_{i=1}^{K} \delta(c, f(x_i)) = 2$$

Logo, x_3 deverá ser classificado como P.

Para x_4:

A K-vizinhança é $\{x_1, x_3, x_5, x_6, x_7\}$.

$$f(x_4) = \underset{c \in Z}{\operatorname{argmax}} \sum_{i=1}^{K} \delta(c, f(x_i))$$

Para $c = P$:

$$\sum_{i=1}^{K} \delta(c, f(x_i)) = 2$$

Para $c = N$:

$$\sum_{i=1}^{K} \delta(c, f(x_i)) = 3$$

Logo, x_4 deverá ser classificado como N.

Para x_5:

A K-vizinhança é $\{x_1, x_2, x_4, x_6, x_8\}$.

$$f(x_5) = \underset{c \in Z}{\operatorname{argmax}} \sum_{i=1}^{K} \delta(c, f(x_i))$$

Para $c = P$:

$$\sum_{i=1}^{K} \delta(c, f(x_i)) = 3$$

Para $c = N$:

$$\sum_{i=1}^{K} \delta(c, f(x_i)) = 2$$

Logo, x_5 deverá ser classificado como P.

Para x_6:

A K-vizinhança é $\{x_1, x_2, x_4, x_5, x_8\}$.

$$f(x_6) = \underset{c \in Z}{\operatorname{argmax}} \sum_{i=1}^{K} \delta(c, f(x_i))$$

Para $c = P$:

$$\sum_{i=1}^{K} \delta(c, f(x_i)) = 3$$

Para $c = N$:

$$\sum_{i=1}^{K} \delta(c, f(x_i)) = 2$$

Logo, x_6 deverá ser classificado como P.

Para x_7:

A k-vizinhança é $\{x_1, x_2, x_3, x_4, x_8\}$.

$$f(x_8) = \underset{c \in Z}{\operatorname{argmax}} \sum_{i=1}^{k} \delta(c, f(x_i))$$

Para $c = P$:

$$\sum_{i=1}^{k} \delta(c, f(x_i)) = 4$$

Para $c = N$:

$$\sum_{i=1}^{k} \delta(c, f(x_i)) = 1$$

Logo, x_7 deverá ser classificado como P.

Para x_8:

A k-vizinhança é $\{x_2, x_3, x_5, x_6, x_7\}$.

$$f(x_8) = \underset{c \in Z}{\operatorname{argmax}} \sum_{i=1}^{k} \delta(c, f(x_i))$$

Para $c = P$:

$$\sum_{i=1}^{k} \delta(c, f(x_i)) = 2$$

Para $c = N$:

$$\sum_{i=1}^{k} \delta(c, f(x_i)) = 3$$

Logo, x_8 deverá ser classificado como N.

Condensando os resultados obtidos numa tabela, temos

	Predicted	True
1	N	P
2	N	P
3	P	P
4	N	P
5	P	N
6	P	N
7	P	N
8	N	N

pred \ true	P	N
P	1	3
N	3	1

$$\text{recall} = \frac{TP}{TP + FN} = \frac{1}{1+3} = 0{,}25$$

$$\text{precision} = \frac{TP}{TP + FP} = \frac{1}{1+3} = 0{,}25$$

$$F1 = \frac{1}{\frac{1}{2}\left(\frac{1}{P} + \frac{1}{R}\right)} = \frac{1}{\frac{1}{2}\left(\frac{1}{0{,}25} + \frac{1}{0{,}25}\right)} = 0{,}25$$

2. Propomos a seguinte métrica:

$$d(x_i, x_j) = \begin{cases} 0, & \text{se } x_{i1} = x_{j1} \\ 1, & \text{caso contrário} \end{cases} \quad \text{e} \quad K = 3$$

Temos então a tabela de distâncias $d(x_i, x_j)$:

i \ j	1	2	3	4	5	6	7	8
1	0	1	0	0	1	1	0	1
2	1	0	1	1	0	0	1	0
3	0	1	0	0	1	1	0	1
4	0	1	0	0	1	1	0	1
5	1	0	1	1	0	0	1	0
6	1	0	1	1	0	0	1	0
7	0	1	0	0	1	1	0	1
8	1	0	1	1	0	0	1	0

$$f(x_k) = \underset{c \in Z}{\operatorname{argmax}} \sum_{i=1}^{K} \delta(c, f(x_i))$$

Para x_1:

A K-vizinhança é $\{x_3, x_4, x_7\}$.

Para $c = P$:

$$\sum_{i=1}^{K} \delta(c, f(x_i)) = 2$$

Para $c = N$:

$$\sum_{i=1}^{K} \delta(c, f(x_i)) = 1$$

Logo, x_1 deve ser classificado como P.

Para x_2:

A K-vizinhança é $\{x_5, x_6, x_8\}$.

Para $c = P$:

$$\sum_{i=1}^{k} S(c, f(x_i)) = 0$$

Para $c = N$:

$$\sum_{i=1}^{K} \delta(c, f(x_i)) = 3$$

Logo, x_2 deverá ser classificado como N.

Para x_3:

A K-vizinhança é $\{x_1, x_4, x_7\}$.

Para $c = P$:

$$\sum_{i=1}^{K} \delta(c, f(x_i)) = 2$$

Para $c = N$:

$$\sum_{i=1}^{K} \delta(c, f(x_i)) = 1$$

Logo, x_3 deverá ser classificado como P.

Para x_4:

A k-vizinhança é $\{x_1, x_3, x_7\}$.

Para $c = P$:

$$\sum_{i=1}^{k} \delta(c, f(x_i)) = 2$$

Para $c = N$:

$$\sum_{i=1}^{k} \delta(c, f(x_i)) = 1$$

Logo, x_4 deverá ser classificado como P.

Para x_5:

A k-vizinhança é $\{x_2, x_6, x_8\}$.

Para $c = P$:

$$\sum_{i=1}^{k} \delta(c, f(x_i)) = 1$$

Para $c = N$:

$$\sum_{i=1}^{k} \delta(c, f(x_i)) = 2$$

Logo, x_5 deverá ser classificado como N.

Para x_6:

A k-vizinhança é $\{x_2, x_5, x_8\}$.

Para $c = P$:

$$\sum_{i=1}^{k} \delta(c, f(x_i)) = 1$$

Para $c = N$:

$$\sum_{i=1}^{k} \delta(c, f(x_i)) = 2$$

Logo, x_6 deverá ser classificado como N.

Para x_7:

A k-vizinhança é $\{x_1, x_3, x_4\}$.

Para $c = P$:

$$\sum_{i=1}^{k} \delta(c, f(x_i)) = 3$$

Para $c = N$:

$$\sum_{i=1}^{k} \delta(c, f(x_i)) = 0$$

Logo, x_7 deverá ser classificado como P.

Para x_8:

A k-vizinhança é $\{x_2, x_5, x_6\}$.

Para $c = P$:

$$\sum_{i=1}^{k} \delta(c, f(x_i)) = 1$$

Para $c = N$:

$$\sum_{i=1}^{k} \delta(c, f(x_i)) = 2$$

Logo, x_8 deverá ser classificado como N.

Condensando os resultados obtidos numa tabela, temos

	Predicted	True
1	P	P
2	N	P
3	P	P
4	P	P
5	N	N
6	N	N
7	P	N
8	N	N

pred \ true	P	N
P	3	1
N	1	3

$$\text{recall} = \frac{TP}{TP + FN} = \frac{3}{3+1} = 0{,}75$$

$$\text{precision} = \frac{TP}{TP + FP} = \frac{3}{3+1} = 0{,}75$$

$$F1 = \frac{1}{\frac{1}{2}\left(\frac{1}{P} + \frac{1}{R}\right)} = \frac{1}{\frac{1}{2}\left(\frac{1}{0{,}75} + \frac{1}{0{,}75}\right)} = 0{,}75$$

Portanto, a medida F1 aumentou em 3 vezes.

3. Para aprender um Bayesian classifier, precisamos de:
- prior probabilities de cada classe (menos uma);
- os parametros associados com $p(D|h)$.

Começemos por calcular as prior probabilities.

$$p(P) = \frac{5}{9} \qquad p(N) = \frac{4}{9}$$

Agora precisamos dos parâmetros associados a $p(D|h)$.

Como y_1 e y_2 são dependentes, temos de calcular a sua probabilidade conjunta.

Para $h = P$:

$y_1 \backslash y_2$	0	1	
A	2/5	1/5	3/5
B	1/5	1/5	2/5
	3/5	2/5	1

Para $h = N$:

$y_1 \backslash y_2$	0	1	
A	0	1/4	1/4
B	2/4	1/4	3/4
	2/4	2/4	1

Como $\{y_1, y_2\}$ e $\{y_3\}$ são conjuntos de variáveis independentes,

$$p(y_1 = a, y_2 = b, y_3 = c \mid h) = P(y_1 = a, y_2 = b \mid h)\, P(y_3 = c \mid h)$$

Logo, precisamos saber $P(y_3 \mid h)$. Como y_3 é normal, precisamos calcular a média e a variância para saber a função densidade de probabilidade.

Para $h = P$:

$$\overline{y_3} = \frac{1}{5}(1{,}1 + 0{,}8 + 0{,}5 + 0{,}9 + 0{,}8) = 0{,}82$$

$$\sigma^2_{y_3} = \frac{1}{4}\left((1{,}1 - \overline{y_3})^2 + (0{,}8 - \overline{y_3})^2 + (0{,}5 - \overline{y_3})^2 + (0{,}9 - \overline{y_3})^2 + (0{,}8 - \overline{y_3})^2\right) = 0{,}047$$

$$y_3 \sim N(\mu = 0{,}82,\ \sigma^2 = 0{,}047)$$

A função de densidade de probabilidade está bem definida.

Para $h = N$:

$$\overline{y_3} = \frac{1}{4}(1 + 0{,}9 + 1{,}2 + 0{,}9) = 1$$

$$\sigma^2_{y_3} = \frac{1}{3}\left((1 - \overline{y_3})^2 + (0{,}9 - \overline{y_3})^2 + (1{,}2 - \overline{y_3})^2 + (0{,}9 - \overline{y_3})^2\right) = 0{,}02$$

$$y_3 \sim N(\mu = 1,\ \sigma^2 = 0{,}02)$$

A função de densidade de probabilidade está bem definida.

4.

$$h_{MAP} = \underset{h}{\text{argmax}} \frac{p(D|h)\,p(h)}{p(D)} = \underset{h}{\text{argmax}}\ p(h) \cdot p(D|h)$$

Para $(A;1;0{,}8)$:

$$h_{MAP} = \underset{h}{\text{argmax}}\ p(h)\, p(y_1 = A,\ y_2 = 1,\ y_3 = 0{,}8 \mid h) =$$

$$= \underset{h}{\text{argmax}}\ p(h)\, p(y_1 = A,\ y_2 = 1 \mid h)\, p(y_3 = 0{,}8 \mid h)$$

Para $h = P$:

$$h_{MAP} = \frac{5}{9} \cdot \frac{1}{5} \cdot \overset{1{,}8324}{p(y_3 = 0{,}8 \mid h = P)} = 0{,}2036$$

Para $h = N$:

$$h_{MAP} = \frac{4}{9} \cdot \frac{1}{4} \cdot \overset{1{,}0378}{p(y_3 = 0{,}8 \mid h = N)} = 0{,}1153$$

Logo, $(A;1;0{,}8)$ deverá ser classificado como P.

Para $(B;1;1)$:

$$h_{MAP} = \underset{h}{\text{argmax}}\ p(h)\, p(y_1 = B,\ y_2 = 1,\ y_3 = 1 \mid h) =$$

$$= \underset{h}{\text{argmax}}\ p(h)\, p(y_1 = B,\ y_2 = 1 \mid h)\, p(y_3 = 1 \mid h)$$

Para $h = P$:

$$h_{MAP} = \frac{5}{9} \cdot \frac{1}{5} \cdot \overset{1{,}3037}{p(y_3 = 1 \mid h = P)} = 0{,}1449$$

Para $h = N$:

$$h_{MAP} = \frac{4}{9} \cdot \frac{1}{4} \cdot \overset{2{,}8210}{p(y_3 = 1 \mid h = N)} = 0{,}3134$$

Logo, $(B;1;1)$ deverá ser classificado como N.

Para $(B; 0; 0{,}9)$:

$$h_{MAP} = \underset{h}{\operatorname{argmax}} \; p(h)\, p(y_1 = B,\, y_2 = 0,\, y_3 = 0{,}9 \mid h) =$$

$$= \underset{h}{\operatorname{argmax}} \; p(h)\, p(y_1 = B,\, y_2 = 0 \mid h)\, p(y_3 = 0{,}9 \mid h)$$

Para $h = P$:

$$h_{MAP} = \frac{5}{9} \cdot \frac{1}{5} \cdot \underbrace{p(y_3 = 0{,}9 \mid h = P)}_{1{,}7191} = 0{,}1910$$

Para $h = N$:

$$h_{MAP} = \frac{4}{9} \cdot \frac{2}{4} \cdot \underbrace{p(y_3 = 0{,}9 \mid h = N)}_{2{,}1970} = 0{,}4882$$

Logo, $(B; 0; 0{,}9)$ deverá ser classificado como N.

5. $h_{ML} = \underset{h}{\operatorname{argmax}} \; p(D \mid h)$