

**TP555 – Inteligência Artificial e Machine Learning**  
**Josino Villela da Silva Neto – Matrícula: 854 – Mestrado em Engenharia de Telecomunicações**

**Lista de Exercícios #4**

1. P – Positivo, D – Doente e S – Saudável

**Dados A:**

$$P(P|D) = 0.95$$

$$P(P|S) = 0.1$$

**Dados B:**

$$P(P|D) = 0.9$$

$$P(P|S) = 0.05$$

$$P(D) = 0.01$$

$$P(S) = 0.99$$

$$P(D|P) = ?$$

**Cálculo para o teste A:**

$$P(D|P) = [P(P|D) \cdot P(D)] / P(P)$$

$$P(D|P) = [P(P|D) \cdot P(D)] / [P(P|D) \cdot P(D) + P(P|S) \cdot P(S)]$$

$$P(D|P) = [0.95 \times 0.01] / [(0.95 \times 0.01) + (0.01 \times 0.99)]$$

$$P(D|P) = 0.0876$$

**Cálculo para o teste B:**

$$P(D|P) = [P(P|D) \cdot P(D)] / P(P)$$

$$P(D|P) = [P(P|D) \cdot P(D)] / [P(P|D) \cdot P(D) + P(P|S) \cdot P(S)]$$

$$P(D|P) = [0.9 \times 0.01] / [(0.9 \times 0.01) + (0.05 \times 0.99)]$$

$$P(D|P) = 0.1548$$

**R:** O teste mais indicativo de alguém realmente estar com o vírus, é o teste B

2. Ca – Casado, Nc – Não casado, Sc – Sem casa, Cc – Com casa, Pg – Pagou e Npg – Não Pagou

$$P(Pg|Sc, Ca, 3) = ?$$

$$P(NPg|Sc, Ca, 3) = ?$$

$$P(Pg|Sc, Ca, 3) = [P(Ca|Pg) \cdot P(Sc|Pg) \cdot P(3|Pg) \cdot P(Pg)] / [P(Ca) \cdot P(Sc) \cdot P(3)]$$

$$P(Pg|Sc, Ca, 3) = [4/7 \times 4/7 \times 2/7 \times 7/10] / [5/10 \times 6/10 \times 3/10]$$

$$P(Pg|Sc, Ca, 3) = 0.7256 \text{ (72.56\%)}$$

$$P(NPg|Sc, Ca, 3) = 1 - 0.7256$$

$$P(NPg|Sc, Ca, 3) = 0.2744 \text{ (27.44\%)}$$

**R:** Se trabalhasse no banco, autorizaria o empréstimo a Jair.

3. Altura (A) – 1.83m, Peso (P) – 58.97 Kg, Calçado (C) – 20.32 cm, M – Masculino, F – Feminino

$$E(A|M) = [1.83 + 1.8 + 1.7 + 1.8] / 4 = 1.7825$$

$$E(P|M) = [81.65 + 86.18 + 77.11 + 74.84] / 4 = 79.945$$

$$E(C|M) = [30.48 + 27.94 + 30.48 + 25.4] / 4 = 28.575$$

$$\text{Var}(A|M) = [(1.83-1.7825)^2 + (1.8-1.7825)^2 + (1.7-1.7825)^2 + (1.8-1.7825)^2] / 4 = 0.03225$$

$$\text{Var}(P|M) = [(81.65-79.945)^2 + (86.18-79.945)^2 + (77.11-79.945)^2 + (74.84-79.945)^2] / 4 = 25.2935$$

$$\text{Var}(C|M) = [(30.48-28.575)^2 + (27.94-28.575)^2 + (30.48-28.575)^2 + (25.4-28.575)^2] / 4 = 5.9139$$

$$E(A|F) = [1.52 + 1.68 + 1.65 + 1.75] / 4 = 1.65$$

$$E(P|F) = [45.36 + 68.04 + 58.97 + 68.04] / 4 = 60.1025$$

$$E(C|F) = [15.24 + 20.32 + 17.78 + 22.86] / 4 = 19.05$$

$$\text{Var}(A|F) = [(1.52-1.65)^2 + (1.68-1.65)^2 + (1.65-1.65)^2 + (1.75-1.65)^2] / 4 = 0.00926$$

$$\text{Var}(P|F) = [(45.36-60.1025)^2 + (68.04-60.1025)^2 + (58.97-60.1025)^2 + (68.04-60.1025)^2] / 4 = 114.8772$$

$$\text{Var}(C|F) = [(15.24-19.05)^2 + (20.32-19.05)^2 + (17.78-19.05)^2 + (22.86-19.05)^2] / 4 = 10.7526$$

$x = A, P \text{ e } C$ , e  $y = M \text{ e } F$

$$P(x|y) = \frac{1}{\text{Var} \cdot \sqrt{2 \cdot \pi}} \cdot e^{\frac{-(x-E)^2}{2 \cdot \text{Var}}}$$

$$P(M) = P(A|M) + P(P|M) + P(C|M) = 3.41E-8$$

$$P(F) = P(A|F) + P(P|F) + P(C|F) = 3E-3$$

Os atributos fornecidos no enunciado têm maior probabilidade de pertencerem ao sexo feminino.

4.

C – Sim. A curva simulada se aproxima da curva teórica da SER.

D – Se tratando de símbolos com diferentes probabilidades, ML não é equivalente a MAP. A equação que maximiza o numerador, é a mesma para ML e MAP, somente se os símbolos são equiprováveis.

5.

A – Para a classe 'comp.os.ms-windows.misc', observa-se que a maioria das predições realizadas estão incorretas.

B – Uma possível causa pela grande quantia de predições erradas, é a utilização de palavras similares na classe 'comp.windows.x'. O que gera confusão no momento da interpretação pelo classificador.

6. As duas mensagens de teste que foram providas no enunciado do exercício, pertencem as classes ham e spam, respectivamente.

7. C – Beijing (Be), Chinese (Ch), Japan (Ja), Macao (Ma), Shanghai (Sh), Tokio (To), China (C) e Not China (NC)

$$P(C) = 0.75 \text{ e } P(NC) = 0.25$$

$$\text{Test} = \{\text{Ch}, \text{Ch}, \text{Ch}, \text{To}, \text{Ja}\}$$

**Bernoulli + suavização de Laplace**

$$X = \{0. \ 1. \ 1. \ 0. \ 0. \ 1\}$$

$$P(\text{Be}|C) = [1+1] / [6+6] = 0.16$$

$$P(\text{Ch}|C) = [3+1] / [6+6] = 0.33$$

$$P(\text{Ja}|C) = [0+1] / [6+6] = 0.083$$

$$P(\text{Ma}|C) = [1+1] / [6+6] = 0.16$$

$$P(\text{Sh}|C) = [1+1] / [6+6] = 0.16$$

$$P(\text{To}|C) = [0+1] / [6+6] = 0.083$$

$$P(\text{Be}|NC) = [0+1] / [3+6] = 0.11$$

$$P(\text{Ch}|NC) = [1+1] / [3+6] = 0.22$$

$$P(\text{Ja}|NC) = [1+1] / [3+6] = 0.22$$

$$P(\text{Ma}|NC) = [0+1] / [3+6] = 0.11$$

$$P(\text{Sh}|NC) = [0+1] / [3+6] = 0.11$$

$$P(\text{To}|NC) = [1+1] / [3+6] = 0.22$$

A priori:

$$P(X|C) = (1-0.16)^3 \times 0.33 \times 0.083 = 0.0013$$

$$P(X|NC) = (1-0.11)^3 \times 0.22^3 = 0.0077$$

A posteriori:

$$P(C|X) = [0.0013 \times 0.75] / 0.003 = 0.34$$

$$P(NC|X) = [0.0077 \times 0.25] / 0.003 = 0.66$$

### Multinomial + suavização de Laplace

$X = \{0. 3. 1. 0. 0. 1\}$

$$P(\text{Be} | C) = [1+1] / [8+6] = 0.14$$

$$P(\text{Ch} | C) = [5+1] / [8+6] = 0.42$$

$$P(\text{Ja} | C) = [0+1] / [8+6] = 0.07$$

$$P(\text{Ma} | C) = [1+1] / [8+6] = 0.14$$

$$P(\text{Sh} | C) = [1+1] / [8+6] = 0.14$$

$$P(\text{To} | C) = [0+1] / [8+6] = 0.07$$

$$P(\text{Be} | \text{NC}) = [0+1] / [3+6] = 0.11$$

$$P(\text{Ch} | \text{NC}) = [1+1] / [3+6] = 0.22$$

$$P(\text{Ja} | \text{NC}) = [1+1] / [3+6] = 0.22$$

$$P(\text{Ma} | \text{NC}) = [0+1] / [3+6] = 0.11$$

$$P(\text{Sh} | \text{NC}) = [0+1] / [3+6] = 0.11$$

$$P(\text{To} | \text{NC}) = [1+1] / [3+6] = 0.22$$

A priori:

$$P(X | C) = [5! / 3! 1! 1!] \times 0.42^3 \times 0.07^2 = 0.0073$$

$$P(X | \text{NC}) = [5! / 3! 1! 1!] \times 0.22^3 \times 0.11^2 = 0.0028$$

A posteriori:

$$P(X) = P(X | C) \cdot P(C) + P(X | \text{NC}) \cdot P(\text{NC})$$

$$P(X) = 0.0073 \times 0.75 + 0.0028 \times 0.25$$

$$P(X) = 0.0062$$

$$P(C | X) = [0.0073 \times 0.75] / 0.0062 = 0.88$$

$$P(\text{NC} | X) = [0.0028 \times 0.25] / 0.0062 = 0.112$$

**D** – O classificador Multinomial considera X um histograma, ou seja, X indica o número de ocorrências de determinado evento. Já o classificador de Bernoulli, considera que X é binário e indica a presença, ou não, de determinado termo (e não o número de ocorrências).