Lista de Exercícios 4 – TP555 Inteligência Artificial e Machine Learning

Aluno: Bruno Ferreira Gomes Matrícula:842

Ex1)

P = Positivo, N = Negativo, D = Doente, S = Saudável.

Dados para A:

$$P(P|D) = 0.95; \qquad P(P|S) = 0.1;$$

Dados para B:

$$P(P|D) = 0.9; \qquad P(P|S) = 0.05;$$

Dado Geral:

$$P(D) = 0.01;$$

Método de cálculo:

$$P(D|P) = \frac{P(P|D)P(D)}{P(P|A)} = \frac{P(P|D)P(D)}{P(P|D)P(D) + P(P|S)P(S)}$$

A:

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

B:

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

Resposta: Observando os valores das probabilidades, podemos concluir que o teste **B** é o melhor, pois assume uma probabilidade menor, sendo que é o teste mais indicativo de que alguém esteja com o vírus.

Ex2)

S = Solteiro, D = Divorciado, Ca = Casado, Pg = Pagou, Npg =Não pagou, PC = Possui casa, NPC = Não possui casa.

$$P(Pg|Ca, NPC, 3) = ?$$
$$P(Npg|Ca, NPC, 3) = ?$$

Dados:

$$P(Ca|Pg) = 4/7$$
$$P(NPC|Pg) = 4/7$$
$$P(3|Pg) = 2/7$$
$$P(Pg) = 0.7$$
$$P(Ca) = 0.5$$
$$P(NPC) = 0.6$$
$$P(3) = 0.3$$

Método de Cálculo:

$$P(Pg|Ca, NPC, 3) = P(Pg|\boldsymbol{X}) = \frac{P(\boldsymbol{X}|Pg)P(Pg)}{P(\boldsymbol{X})}$$

$$= \frac{P(Ca|Pg)P(NPC|Pg)P(3|Pg)P(Pg)}{P(Ca)P(NPC)P(3)}$$

Assim:

$$P(Pg|Ca, NPC, 3) = 0.7256\ (72.6\%)$$

E então:

$$P(Npg|Ca, NPC, 3) = 1 - P(Pg|Ca, NPC, 3) =$$
$$P(Npg|Ca, NPC, 3) = 0.2744\ (27.4\%)$$

<u>**Resposta:**</u> Devido aos valores de probabilidade adquiridos, eu aprovaria o empréstimo ao Jair.

Ex3)

Altura (A) = 1.83, Peso (P) = 58.97 kg, Tamanho Calçado (TC) = 20.32 cm, M = Masculino, F = Feminino.

Assim:

$$E[A|M] = [1.83 + 1.8 + 1.7 + 1.8]/4 = 1.7825$$
$$E[P|M] = [81.35 + 86.18 + 77.11 + 74.84]/4 = 79.945$$
$$E[TC|M] = [30.48 + 27.94 + 30.48 + 25.4]/4 = 28.575$$

$$Var[A|M] = [(1.83 - 1.7825)^2 + (1.8 - 1.7825)^2 + (1.7 - 1.7825)^2 + (1.8 - 1.7825)^2]/4 = 0.03225$$
$$Var[P|M] = [(81.35 - 79.945)^2 + (86.18 - 79.945)^2 + (77.11 - 79.945)^2 + (74.84 - 79.945)^2]/4 = 25.2935$$
$$Var[TC|M] = [(30.48 - 28.575)^2 + (27.94 - 28.575)^2 + (30.48 - 28.575)^2 + (25.4 - 28.575)^2]/4 = 5.9139$$

$$E[A|F] = [1.52 + 1.68 + 1.65 + 1.75]/4 = 1.65$$
$$E[P|F] = [45.36 + 68.04 + 58.97 + 68.04]/4 = 60.1025$$
$$E[TC|F] = [15.24 + 20.32 + 17.78 + 22.86]/4 = 19.05$$

$$Var[A|F] = [(1.52 - 1.65)^2 + (1.68 - 1.65)^2 + (1.65 - 1.65)^2 + (1.75 - 1.65)^2]/4 = 0.00926$$
$$Var[P|F] = [(45.36 - 60.1025)^2 + (68.04 - 60.1025)^2 + (58.97 - 60.1025)^2 + (68.04 - 60.1025)^2]/4 = 114.8772$$
$$Var[TC|M] = [(15.24 - 19.05)^2 + (20.32 - 19.05)^2 + (17.78 - 19.05)^2 + (22.86 - 19.05)^2]/4 = 10.7526$$

Sendo X = A, P, TC e Y = M, F

Temos:

Método de Cálculo:

$$P(x|Y) = \frac{1}{Var[X|Y] * \sqrt{2\,\pi}} exp\left(-\frac{(x - E[X|Y])^2}{2 * Var[X|Y]^2}\right)$$

P(M) = P(A|M) + P(P|M) + P(TC|M) = 3.41e-8

P(F) = P(A|F) + P(P|F) + P(TC|F) = 3e-3

Resposta: Os dados apresentados têm mais chance de ser do sexo feminino, devido ao fato de ter uma probabilidade maior.

Ex7)

Letra C)

Be = Beijing, Ch = Chinese, Ja= Japan, Ma = Macao, Sh = Shanghai, To = Tokyo, C = China, NC = Not China.

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

$P(C|\boldsymbol{X}) = \frac{P(\boldsymbol{X}|C)\, P(C)}{P(\boldsymbol{X})} =?$

$P(NC|\boldsymbol{X}) = \frac{P(\boldsymbol{X}|NC)\, P(NC)}{P(\boldsymbol{X})} = ?$

Multinominal + suavização de Laplace

$$\boldsymbol{X} = \{\, \mathbf{0 \;\; 3 \;\; 1 \;\; 0 \;\; 0 \;\; 1} \,\}$$

$$P(Be|C) = \frac{1+1}{8+6} = 0.14$$

$$P(Ch|C) = \frac{5+1}{8+6} = 0.42$$

$$P(Ja|C) = \frac{0+1}{8+6} = 0.07$$

$$P(Ma|C) = \frac{1+1}{8+6} = 0.14$$

$$P(Sh|C) = \frac{1+1}{8+6} = 0.14$$

$$P(To|C) = \frac{0+1}{8+6} = 0.07$$

xxxxxxxxxxxxxxxxxxxxxxxxxxxxxx

$$P(Be|NC) = \frac{0+1}{3+6} = 0.11$$

$$P(Ch|NC) = \frac{1+1}{3+6} = 0.22$$

$$P(Ja|NC) = \frac{1+1}{3+6} = 0.22$$

$$P(Ma|NC) = \frac{0+1}{3+6} = 0.11$$

$$P(Sh|NC) = \frac{0+1}{3+6} = 0.11$$

$$P(To|NC) = \frac{1+1}{3+6} = 0.22$$

$P(C) = 3/4$ e $P(NC) = 1/4$

A priori:

$$P(\boldsymbol{X}|C) = \frac{5!}{3!\ 1!\ 1!} 0.42^3 * 0.07 * 0.07 = 0.0073$$
$$P(\boldsymbol{X}|NC) = \frac{5!}{3!\,1!\ 1!} 0.222^3 * 0.222 * 0.222 = 0.0107$$

A posteriori:

$$P(\boldsymbol{X}) = P(\boldsymbol{X}|C)P(C) + P(\boldsymbol{X}|NC)P(NC) = 0.00812$$

$$P(C|\boldsymbol{X}) = \frac{0.00726 * 0.75}{0.00812} = 0.67$$

$$P(NC|\boldsymbol{X}) = \frac{0.0107 * 0.25}{0.00812} = 0.33$$

Bernoulli + suavização de Laplace

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

$$P(Be|C) = \frac{1+1}{6+6} = 0.16$$

$$P(Ch|C) = \frac{3+1}{6+6} = 0.33$$

$$P(Ja|C) = \frac{0+1}{6+6} = 0.083$$

$$P(Ma|C) = \frac{1+1}{6+6} = 0.16$$

$$P(Sh|Ch) = \frac{1+1}{6+6} = 0.16$$

$$P(To|C) = \frac{0+1}{6+6} = 0.083$$

xxxxxxxxxxxxxxxxxxxxxxxxxxxxx

$$P(Be|NC) = \frac{0+1}{3+6} = 0.11$$

$$P(Ch|NC) = \frac{1+1}{3+6} = 0.22$$

$$P(Ja|NC) = \frac{1+1}{3+6} = 0.22$$

$$P(Ma|NC) = \frac{0+1}{3+6} = 0.11$$

$$P(Sh|NC) = \frac{0+1}{3+6} = 0.11$$

$$P(To|NC) = \frac{1+1}{3+6} = 0.22$$

$P(C) = 3/4$ e $P(NC) = 1/4$

A priori

$$P(\boldsymbol{X}|C) = (1-0.166)^3 * 0.333 * 0.083^2 = 0.00133$$
$$P(\boldsymbol{X}|NC) = (1-0.111)^3 * 0.222^3 = 0.00769$$

A posteriori

$$P(\boldsymbol{X}) = P(\boldsymbol{X}|C)P(C) + P(\boldsymbol{X}|NC)P(NC) = 0.00291$$

$$P(C|\boldsymbol{X}) = \frac{0.00133 * 0.75}{0.00291} = 0.34$$

$$P(NC|\boldsymbol{X}) = \frac{0.00769 * 0.25}{0.00291} = 0.66$$

Resposta Letra D: Ao se utilizar o classificador de multinominal, este leva em conta a quantidade de ocorrências de um evento, pois X denota um histograma, composto por eventos distintos. No classificador de Bernoulli, é diferente, agindo de forma oposta ao multinominal, onde no caso, este considera menos valores para X, renunciando à quantidade de ocorrências de em evento, por ser considerado como uma escolha binária, e reduzindo no caso para a escolha de um termo.