

Lista #4

Exercício 1

Neste exercício você utilizará o teorema de Bayes. Considere dois exames médicos, A e B, para um vírus. O teste A é 95% eficaz no reconhecimento do vírus quando ele está presente, mas tem uma taxa de falso positivo de 10% (indicando que o vírus está presente, quando ele não está). O teste B é 90% eficaz no reconhecimento do vírus, mas possui uma taxa de falso positivo de 5%. Os dois testes usam métodos independentes para identificar o vírus. 1% de todas as pessoas possuem o vírus. Digamos que uma pessoa é testada para o vírus usando apenas um dos testes e que o teste é positivo para o vírus. Qual teste, retornando positivo, é mais indicativo de alguém realmente estar com o vírus?

Classes	Atributos
Tem o vírus	Positivo
Não tem o vírus	Negativo

Exame Médico A

$$P(+|vírus) = 0.95$$

$$P_{vírus} = 0.01$$

$$P(+|sem_vírus) = 0.1$$

$$P_{sem_vírus} = 0.99$$

$$P_{(+)} = P_{(+|vírus)}P_{vírus} + P_{(+|sem_vírus)}P_{sem_vírus} = 0.1085$$

$$P(vírus|+) = \frac{P(+|vírus)P(vírus)}{P(+)} = 87.5576e^{-3}$$

$$P(sem_vírus|+) = \frac{P(+|sem_vírus)P(sem_vírus)}{P(+)} = 912.4423e^{-3}$$

Exame Médico B

$$P(+|vírus) = 0.9$$

$$P_{vírus} = 0.01$$

$$P(+|sem_vírus) = 0.05$$

$$P_{sem_vírus} = 0.99$$

$$P_{(+)} = P_{(+|vírus)}P_{vírus} + P_{(+|sem_vírus)}P_{sem_vírus} = 0.0585$$

$$P(vírus|+) = \frac{P(+|vírus)P(vírus)}{P(+)} = 153.8461e^{-3}$$

$$P(sem_vírus|+) = \frac{P(+|sem_vírus)P(sem_vírus)}{P(+)} = 846.1538e^{-3}$$

O método B apresenta um indicativo maior de que alguém esteja com o vírus.

Exercício 3

Neste exercício você vai prever, baseado em alguns atributos físicos de uma pessoa, se ela é do sexo masculino ou feminino. Dado os seguintes atributos físicos de uma pessoa: altura = 1.83 metros, peso = 58.97 Quilos e tamanho do calçado = 20.32 centímetros. Baseado nas informações anteriores, qual classe tem maior probabilidade, ou seja, qual dos 2 sexos teria a maior probabilidade? Para calcular as probabilidades, utilize os dados da tabela abaixo. **Obs.:** Apresente todos os cálculos feitos para se encontrar as probabilidades de cada classe, ou seja, neste exercício você não deve utilizar a biblioteca SciKit-learn.

(Dica: Assuma que os atributos seguem uma distribuição Gaussiana).

(Dica: Assuma que a probabilidade da pessoa ser do sexo masculino ou do feminino é de 0.5, respectivamente).

(Dica: utilize a teoria do classificador naive Bayes e lembre-se que o numerador da equação do classificador não influencia na maximização das probabilidades).

Altura [m]	Peso [kg]	Tamanho calçado [cm]	Sexo
1.83	81.65	30.48	Masculino
1.80	86.18	27.94	Masculino
1.70	77.11	30.48	Masculino
1.80	74.84	25.40	Masculino
1.52	45.36	15.24	Feminino
1.68	68.04	20.32	Feminino
1.65	58.97	17.78	Feminino
1.75	68.04	22.86	Feminino

Fórmulas:

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i$$

$$\sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

$$P(x_k | C_q) = \frac{1}{\sigma^2_{x_k, C_q} \sqrt{2\pi}} e^{-\frac{(x_k - \mu_{x_k, C_q})^2}{2\sigma^2_{x_k, C_q}}}$$

$$P_{\text{masc}} | 1.83, 58.97, 20.32 = P_{1.83 | \text{masc}} * P_{58.97 | \text{masc}} * P_{20.32 | \text{masc}} * P_{\text{masc}}$$

$$P_{\text{fem}} | 1.83, 58.97, 20.32 = P_{1.83 | \text{fem}} * P_{58.97 | \text{fem}} * P_{20.32 | \text{fem}} * P_{\text{fem}}$$

Resolução:

$$P_{masc} = \frac{4}{8} = 0.5$$

$$P_{fem} = \frac{4}{8} = 0.5$$

$$\mu_{alt_masc} = \frac{1.83 + 1.80 + 1.70 + 1.80}{4} = 1.7825$$

$$\sigma_{alt_masc} = \sqrt{\frac{[(1.83 - 1.7825)^2 + (1.80 - 1.7825)^2 + (1.70 - 1.7825)^2 + (1.80 - 1.7825)^2]}{4}} = 0.04918078893$$

$$\mu_{alt_fem} = \frac{1.52 + 1.68 + 1.65 + 1.75}{4} = 1.65$$

$$\sigma_{alt_fem} = \sqrt{\frac{[(1.52 - 1.65)^2 + (1.68 - 1.65)^2 + (1.65 - 1.65)^2 + (1.75 - 1.65)^2]}{4}} = 0.08336666$$

$$\mu_{peso_masc} = \frac{81.65 + 86.18 + 77.11 + 74.84}{4} = 79.945$$

$$\sigma_{peso_masc} = \sqrt{\frac{[(81.65 - 79.945)^2 + (86.18 - 79.945)^2 + (77.11 - 79.945)^2 + (74.84 - 79.945)^2]}{4}} = 4.355470698$$

$$\mu_{alt_fem} = \frac{45.36 + 68.04 + 58.97 + 68.04}{4} = 60.1025$$

$$\sigma_{peso_fem} = \sqrt{\frac{[(45.36 - 60.1025)^2 + (68.04 - 60.1025)^2 + (58.97 - 60.1025)^2 + (68.04 - 60.1025)^2]}{4}} = 9.282128999$$

$$\mu_{cal_masc} = \frac{30.48 + 27.94 + 30.48 + 25.40}{4} = 28.575$$

$$\sigma_{cal_masc} = \sqrt{\frac{[(30.48 - 28.575)^2 + (27.94 - 28.575)^2 + (30.48 - 28.575)^2 + (25.40 - 28.575)^2]}{4}} = 2.106056742$$

$$\mu_{cal_fem} = \frac{15.24 + 20.32 + 17.78 + 22.86}{4} = 19.05$$

$$\sigma_{cal_fem} = \sqrt{\frac{[(15.24 - 19.05)^2 + (20.32 - 19.05)^2 + (17.78 - 19.05)^2 + (22.86 - 19.05)^2]}{4}} = 2.839806331$$

$$P_{1.83 | masc} = \frac{1}{0.04918078893\sqrt{2\pi}} e^{-\frac{(1.83 - 1.7825)^2}{2(0.04918078893)^2}} = 5.088104675$$

$$P_{1.83 | fem} = \frac{1}{0.08336666\sqrt{2\pi}} e^{-\frac{(1.83 - 1.65)^2}{2(0.08336666)^2}} = 0.4651631962$$

$$P_{58.97 | masc} = \frac{1}{4.355470698\sqrt{2\pi}} e^{-\frac{(58.97 - 79.945)^2}{2(4.355470698)^2}} = 8.43039314777e^{-7}$$

$$P_{58.97 | fem} = \frac{1}{9.282128999\sqrt{2\pi}} e^{-\frac{(58.97 - 60.1025)^2}{2(9.282128999)^2}} = 0.0426608975$$

$$P_{20.32 | masc} = \frac{1}{2.106056742\sqrt{2\pi}} e^{-\frac{(20.32 - 28.575)^2}{2(2.106056742)^2}} = 8.735117487e^{-5}$$

$$P_{20.32 | fem} = \frac{1}{2.839806331\sqrt{2\pi}} e^{-\frac{(20.32 - 19.05)^2}{2(2.839806331)^2}} = 0.1271135637$$

$$P_{masc | 1.83, 58.97, 20.32} = 5.088104675 * 8.430393148 * 8.735117487 * 0.5 = 1.8734522154e^{-10}$$

$$P_{fem | 1.83, 58.97, 20.32} = 0.4651631962 * 0.0426608975 * 0.1271135637 * 0.5 = 1.261238539e^{-3}$$

A probabilidade é maior para ser do sexo **feminino**.