Quiz 2 Solutions

CSE 4/574

September 18, 2024

Question 1

Day	Outlook	Temperature	Humidity	Wind	PlayTennis
D1	Sunny	Hot	High	Weak	No
D2	Sunny	Hot	High	Strong	No
D3	Overcast	Hot	High	Weak	Yes
D4	Rain	Mild	High	Weak	Yes
D5	Rain	Cool	Normal	Weak	Yes
D6	Rain	Cool	Normal	Strong	No
D7	Overcast	Cool	Normal	Strong	Yes
D8	Sunny	Mild	High	Weak	No
D9	Sunny	Cool	Normal	Weak	Yes
D10	Rain	Mild	Normal	Weak	Yes
D11	Sunny	Mild	Normal	Strong	Yes
D12	Overcast	Mild	High	Strong	Yes
D13	Overcast	Hot	Normal	Weak	Yes
D14	Rain	Mild	High	Strong	No

Table 1: The play tennis dataset

Assume that PlayTennis is the binary target class. Select the correct statement with respect to Naive Bayes Classification.

Correct Choice

For a new instance, x = (Outlook = Sunny, Temperature = Cool, Humidity = High, Wind = Weak), the class conditional probability $P(x \mid PlayTennis = No)$ is $\frac{24}{625}$.

Choice Explanation:

$$
\begin{aligned}
&P(x \mid PlayTennis = No) \\
&= P(Sunny \mid No) * P(Cool \mid No) * P(High \mid No) * P(Weak \mid No) \\
&= \frac{3}{5} * \frac{1}{5} * \frac{4}{5} * \frac{2}{5} \\
&= \frac{24}{625}
\end{aligned}
$$

Incorrect Choice 1

For a new instance, x = (Outlook = Sunny, Temperature = Cool, Humidity = High, Wind = Weak), the class conditional probability for class Yes = $P(x \mid PlayTennis = Yes)$ is $\frac{2}{243}$.

Choice Explanation:

$$
\begin{aligned}
&P(x \mid PlayTennis = Yes) \\
&= P(Sunny \mid Yes) * P(Cool \mid Yes) * P(High \mid Yes) * P(Weak \mid Yes) \\
&= \frac{2}{9} * \frac{3}{9} * \frac{3}{9} * \frac{6}{9} \\
&= \frac{2}{9} * \frac{1}{3} * \frac{1}{3} * \frac{2}{3} \\
&= \frac{4}{243}
\end{aligned}
\tag{1}
$$

Incorrect Choice 2

For a new instance, x = (Outlook = Sunny, Temperature = Cool, Humidity = High, Wind = Weak), the class conditional probability $P(x \mid PlayTennis = No)$ is $\frac{239}{243}$.

Choice Explanation:
See Explanation for the Correct Choice.

Incorrect Choice 3

Probability $P(PlayTennis = Yes)$ is 0.5, assuming no prior on the distribution.

Choice Explanation:
For 9 out of 14 days, playTennis = Yes. Therefore, $P(PlayTennis = Yes) = \frac{9}{14}$

Question 2

Suppose that we have no way of determining if a given fruit is an apple or an orange except by measuring its weight. We have estimated that $\mu_{apples} = 3$, $\mu_{oranges} = 5$, and $\sigma_{apples} = 0.5$, $\sigma_{oranges} = 2$.

Suppose now that we are given an unknown fruit which weighs 4 units. Choose the correct option among the following:

Hint: Compute likelihood of the fruit belonging to any mixture component using the corresponding normal distribution, where k belongs apples or oranges. Then use Bayes' rule to compute the posterior probability, $p(z = k \mid x = 4)$.

$$p(x = 4|z = k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\left(-\frac{1}{2\sigma_k^2}(x - \mu_k)^2\right)$$

Correct Choice

It is more likely that the given fruit is an orange when the priors are uniform.

Choice Explanation:

$$\begin{aligned}
&p(z = apple \mid x = 4) \\
&= p(x = 4 \mid z = apple) \cdot p(z = apple) \\
&= \frac{1}{\sqrt{\frac{1}{2}\pi}} e^{-\frac{1}{0.5}(4-3)^2} \cdot p(z = apple) \\
&= \frac{1}{\sqrt{\frac{1}{2}\pi}} e^{-2} \cdot p(z = apple) \\
&= 0.108 \cdot p(z = apple)
\end{aligned}$$

$$
\begin{aligned}
&p(z = orange \mid x = 4) \\
&= p(x = 4 \mid z = orange) \cdot p(z = orange) \\
&= \frac{1}{\sqrt{8\pi}} e^{-\frac{1}{8}(4-5)^2} \cdot p(z = orange) \\
&= \frac{1}{\sqrt{8\pi}} e^{-\frac{1}{8}} \cdot p(z = orange) \\
&= 0.176 \cdot p(z = orange)
\end{aligned}
$$

Given uniformed prior, $p(z = apple) = p(z = orange)$. Therefore, $p(z = apple \mid x = 4) < p(z = orange \mid x = 4)$.

Incorrect Choice 1

It is more likely that the given fruit is an orange, for any prior probability for the z.

Choice Explanation:
As a counter example, for prior distribution with $p(z = apple) = 0.9$, $p(z = orange) = 0.1$,

$$
\begin{aligned}
p(z = apple \mid x = 4) &= 0.108 * 0.9 = 0.0972 \\
p(z = orange \mid x = 4) &= 0.176 * 0.1 = 0.0176
\end{aligned}
$$

Incorrect Choice 2

It is more likely that the given fruit is an orange when the prior probability of being an apple is known to be $\frac{2}{3}$ and being an orange is known to be $\frac{1}{3}$.

Choice Explanation:

$$
\begin{aligned}
&p(z = apple \mid x = 4) \\
&= p(x = 4 \mid z = apple) \cdot p(z = apple) \\
&= 0.108 \cdot \frac{2}{3} \\
&= 0.072
\end{aligned}
$$

$$
\begin{aligned}
&p(z = orange \mid x = 4) \\
&= p(x = 4 \mid z = orange) \cdot p(z = orange) \\
&= 0.176 \cdot \frac{1}{3} \\
&= 0.059
\end{aligned}
$$

Incorrect Choice 3

The given data is insufficient to answer this question.

Choice Explanation:
Refer to explanation above.