Gabriel Buginga

# Machine Learning
# CPS 863
Lista 1

September 29, 2019

| | |
|---|---|
| Date Performed: | September 29, 2019 |
| Institution: | PESC/COPPE/UFRJ |
| Instructors: | Edmundo de Souza e Silva |
| | Rosa M. M. Leão |
| | Daniel Sadoc Menasché |

## Problem 1

Let $X_{doente}$, $teste\_positivo$, $teste\_negativo$, $tem\_doenca$, $nao\_doenca$ be the obvious relevant events from the problem definition. We explicitly know $P(X_{doente}) = 0.001$, $P(teste\_positivo|tem\_doenca) = 0.9$, $P(teste\_negativo|nao\_doenca) = 0.8$; and we want $P(X_{doente}|teste\_positivo)$. In order to do that the target need to be rewritten with Bayes's theorem using the binary event decomposition of having or not the illness[1]:

$$\begin{aligned} P(X_{doente}|teste\_positivo) &= \frac{P(X_{doente})P(teste\_positivo|X_{doente})}{P(X_{doente})P(teste\_positivo|X_{doente}) + P(\overline{X_{doente}})P(teste\_positivo|\overline{X_{doente}})} \\ &= \frac{0.001 \cdot 0.9}{0.001 \cdot 0.9 + (1 - 0.001) \cdot (1 - 0.8)} \\ &= 0.004484 \end{aligned}$$

## Problem 2

First let's translate the problem information to probability events; to make writing easier $T = trafego$, $\overline{T} = sem\_trafego$, $C = chovendo$, $\overline{C} = sem\_chuva$, $M_{late} = maria\_atrasada$, $\overline{M_{late}} = maria\_pontual$. Information given: $P(C) = \frac{1}{3}$, $P(T|C) = \frac{1}{2}$, $P(T|\overline{C}) = \frac{1}{4}$, $P(M_{late}|C \cap T) = \frac{1}{2}$, $P(M_{late}|\overline{C} \cap \overline{T}) = \frac{1}{8}$, $P(M_{late}|C \cap \overline{T}) = \frac{1}{4}$, $P(M_{late}|\overline{C} \cap T) = \frac{1}{4}$.

**1**: We clearly want $P(\overline{M_{late}} \cap \overline{C} \cap \overline{T})$; using conditional independence[1]:

$$\begin{aligned}
P(\overline{M_{late}} \cap \overline{C} \cap \overline{T}) &= P(\overline{M_{late}}|\overline{C} \cap \overline{T}) \cdot P(\overline{C} \cap \overline{T}) \\
&= P(\overline{M_{late}}|\overline{C} \cap \overline{T}) \cdot P(\overline{T}|\overline{C}) \cdot P(\overline{C}) \\
&= \left(1 - \frac{1}{8}\right) \cdot \left(1 - \frac{1}{4}\right) \cdot \left(1 - \frac{1}{3}\right) \\
&= 0.4375
\end{aligned}$$

**2**: Let's use Bayes' theorem with $M_{late}$, then we just complement:

$$\begin{aligned}
P(M_{late}) &= P(M_{late}|\overline{C} \cap \overline{T}) \cdot P(\overline{C} \cap \overline{T}) + P(M_{late}|\overline{C} \cap T) \cdot P(\overline{C} \cap T) \\
&\quad + P(M_{late}|C \cap \overline{T}) \cdot P(C \cap \overline{T}) + P(M_{late}|C \cap T) \cdot P(C \cap T) \\
&= P(M_{late}|\overline{C} \cap \overline{T}) \cdot P(\overline{T}|\overline{C}) \cdot P(\overline{C}) + P(M_{late}|\overline{C} \cap T) \cdot P(T|\overline{C}) \cdot P(\overline{C}) \\
&\quad + P(M_{late}|C \cap \overline{T}) \cdot P(\overline{T}|C) \cdot P(C) + P(M_{late}|C \cap T) \cdot P(T|C) \cdot P(C) \\
&= \left(\frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{3}\right) + \left(\frac{1}{4} \cdot \frac{1}{2} \cdot \frac{1}{3}\right) + \left(\frac{1}{4} \cdot \frac{1}{4} \cdot \frac{2}{3}\right) + \left(\frac{1}{8} \cdot \frac{3}{4} \cdot \frac{2}{3}\right) \\
&= 0.22916 \\
P(\overline{M_{late}}) &= 1 - P(M_{late}) \\
&= 0.7708
\end{aligned}$$

**3**: Formulating the problem's requirement with events, we want: $P(C|M_{late})$

$$\begin{aligned}
P(C|M_{late}) &= \frac{P(M_{late} \cap C)}{M_{late}} \\
&= \frac{P(M_{late} \cap C \cap T) + P(M_{late} \cap C \cap \overline{T})}{M_{late}} \\
&= \frac{P(M_{late}|C \cap \overline{T}) \cdot P(\overline{T}|C) \cdot P(C) + P(M_{late}|C \cap T) \cdot P(T|C) \cdot P(C)}{M_{late}} \\
&= \frac{\left(\frac{1}{4} \cdot \frac{1}{4} \cdot \frac{2}{3}\right) + \left(\frac{1}{8} \cdot \frac{3}{4} \cdot \frac{2}{3}\right)}{0.22916} \\
&= 0.4545
\end{aligned}$$

## Problem 3

Events are defined as $H = heads$, $T = tails$, $N_1, N_2 = nonbiased\_coins$ and $B = biased\_coin$. Then, directly from the problem definition we have: $P(N_1) = P(N_2) = P(B) = \frac{1}{3}$, $P(H|B) = 1$ and the obvious rest. For $P(H)$ we use the formula for total probability[1]:

$$\begin{aligned} P(H) &= P(H|N_1)P(N_1) + P(H|N_2)P(N_2) + P(H|B)P(B) \\ &= \left(\frac{1}{2} \cdot \frac{1}{3}\right) + \left(\frac{1}{2} \cdot \frac{1}{3}\right) + \left(1 \cdot \frac{1}{3}\right) \\ &= 0.666\overline{6} \end{aligned}$$

For the other answer, a Bayes' formula suffices:

$$\begin{aligned} P(B|H) &= \frac{P(H|B)P(B)}{P(H)} \\ &= \frac{1 \cdot \frac{1}{3}}{0.666\overline{6}} \\ &= 0.5000 \end{aligned}$$

## Problem 4

We have essentially three events: $A$ and $B$ models the corresponding tests returning positive, $\overline{A}$ and $\overline{B}$ negative; $I$ and $\overline{I}$ if a member is in fact ill or not. Data from problem's statement: $P(I) = 0.01$, $P(\overline{A}|I) = 0.1$, $P(\overline{B}|I) = 0.05$, $P(A|\overline{I}) = 0.06$ and $P(B|\overline{I}) = 0.08$.

**1**: We want $P(I|A \cap B)$; using the hypotheses of conditional independence of $A$ and $B$ given $C$, i.e.$(A \perp B)|I$, and taking note that a priori $P(A \cap B)$ is not necessarily equal to $P(A) \cdot P(B)$. Then the following relation can be derived[1]:

$$\begin{aligned} P(I|A \cap B) &= \frac{P(A \cap B|I) \cdot P(I)}{P(A \cap B)} \\ &= \frac{P(A|I) \cdot P(B|I) \cdot P(I)}{P(A \cap B|I) \cdot P(I) + P(A \cap B|\overline{I}) \cdot P(\overline{I})} \\ &= \frac{P(A|I) \cdot P(B|I) \cdot P(I)}{P(A|I) \cdot P(B|I) \cdot P(I) + P(A|\overline{I}) \cdot P(B|\overline{I}) \cdot P(\overline{I})} \\ &= 0.6427 \end{aligned}$$

**2**: Aiming for $P(B|A)$ and using similar equalities from above:

$$\begin{aligned}
P(B|A) &= \frac{P(A\cap B)}{P(A)} \\
&= \frac{P(A\cap B|I)\cdot P(I) + P(A\cap B|\overline{I})\cdot P(\overline{I})}{P(A)} \\
&= \frac{P(A|I)\cdot P(B|I)\cdot P(I) + P(A|\overline{I})\cdot P(B|\overline{I})\cdot P(\overline{I})}{P(A|I)\cdot P(I) + P(A|\overline{I})\cdot P(\overline{I})} \\
&= 0.1944
\end{aligned}$$

**3**: Finally, for our needed $P(A\cup B|I)$, we use two Bayes and the probability of the union:

$$\begin{aligned}
P(A\cup B|I)) &= \frac{P((A\cup B)\cap I)}{P(I)} \\
&= \frac{P((A\cap I)\cup(B\cap I))}{P(I)} \\
&= \frac{P(A\cap I) + P(B\cap I) - P(A\cap B\cap I)}{P(I)} \\
&= \frac{P(A|I)\cdot P(I) + P(B|I)\cdot P(I) - P(A|I)\cdot P(B|I)\cdot P(I)}{P(I)} \\
&= P(A|I) + P(B|I) - P(A|I)\cdot P(B|I) \\
&= 0.9950
\end{aligned}$$

# References

[1] A. N. Shiryaev, "Probability-1," 2016.