

Druhé cvičení - řešení

Matej Lieskovský

Podmíněné pravděpodobnosti

Předpokládejme $P(A) > 0$ a $P(B) > 0$

$$\begin{aligned} P(A|B) > P(A) &\iff P(A \cap B)/P(B) > P(A) \iff P(A \cap B) > P(A)P(B) \iff \\ &\iff P(A \cap B)/P(A) > P(B) \iff P(B|A) > P(B) \end{aligned}$$

Tvrzení jsou tedy ekvivalentní pokud jsou pravděpodobnosti nenulové.

Zkusme si ale rozmyslet, co se stane, pokud alespoň jedna pravděpodobnost je nulová. Budeme ale muset postupovat velmi opatrně.

Pokud je $P(A) = 0$ a současně $P(A|B) > 0$, tak víme, že $P(B) = 0$, protože jinak $P(A) \geq P(A \cap B) = P(A|B)P(B) > 0$. Takže $P(A) = 0$ a $P(B) = 0$.

Pokud by tedy alespoň jedna z těchto pravděpodobností byla nula, tak jsou nulové obě, a tak obě tvrzení jsou nedefinované. Proč přesně nemůžeme definovat podmíněnou pravděpodobnost pro podmínky s nulovou pravděpodobností si ukážeme příště.

Nezávislost

1

Potřebujeme následující pozorování:

$$P(B^C) = 1 - P(B)$$

$$P(A) = P(A \cap B) + P(A \cap B^C)$$

Následně je celý důkaz jednoduchý

$$P(A)P(B^C) = P(A)(1 - P(B)) = P(A) - P(A)P(B) = P(A) - P(A \cap B) = P(A \cap B^C)$$

Nezávislost A^C a B^C pak dokážeme opakováním výše uvedeného procesu.

2

Hodíme korunou a dvojkorunou.

A = na koruně padl orel B = na dvojkoruně padl orel C = obě mince padly stejně

3

Vyrobíme tři jevy A, B, C každý s pravděpodobností $9/27 = 1/3$. Průnik všech tří bude mít pravděpodobnost $1/3^3 = 1/27$, ale průnik libovolných dvou bude mít pravděpodobnost $5/27 \neq 1/3^2$

Existence takového systému lze nahlédnout Vennovým diagramem. Pro jistotu ale můžeme vyrobit 27 míček s čísly 30,6,12,18,24,10,20,40,50,15,45,75,105,1,7,11,13,17,19,23,29,31,37,41,43,47,53, vytáhnout náhodný míček a jako ty tři jevy používat dělitelnost 2, 3 a 5.

Spam filter

Nechť S značí to, že mail je spam, a O značí to, že mail je označen filtrem.

Víme že:

- $P(S) = 0.8$
- $P(O|S) = 0.9$

- $P(O|S^C) = 0.05$

Otázky jsou následující:

$$P(O) = P(O \wedge S) + P(O \wedge S^C) = P(O|S)P(S) + P(O|S^C)P(S^C) = 0.9 \times 0.8 + 0.05 \times 0.2 = 0.72 + 0.01 = 0.73$$

$$P(S^C|O) = P(S^C \wedge O)/P(O) = 0.01/0.73 = 0.01369\dots$$

$$P(S|O^C) = P(S \wedge O^C)/P(O^C) = (P(S) - P(O \wedge S))/(1 - P(O)) = (0.8 - 0.72)/(1 - 0.73) = 0.08/0.27 = 0.296\dots$$

Obrázky

Nechť P značí to, že obrázek obsahuje psa, a O značí, že byl algoritmem tak označen. Z výsledků pro první a druhou skupinu víme že:

- $P(O|P) = 0.8$
- $P(O|P^C) = 0.1$

Pro otázky použijeme následující postup:

Do vzorce $P(O) = P(O|P)P(P) + P(O|P^C)(1 - P(P))$ dosadíme známá čísla a vyjádříme $P(P) = (P(O) - 0.1)/0.7$, čímž dostaneme počet psů. Následně spočítáme $P(P|O) = P(P \wedge O)/P(O) = P(O|P)P(P)/P(O)$ a $P(P|O^C) = P(P \wedge O^C)/P(O^C) = P(O^C|P)P(P)/P(O^C)$.

Ve třetí skupině je tedy asi 429 obrázků psů a pravděpodobnosti jsou 0.857 a 0.142. Ve čtvrté skupině je asi 214 obrázků psů a pravděpodobnosti jsou 0.686 a 0.057.

Všimněte si, že pravděpodobnost toho, že označený obrázek je obrázek psa, se mezi skupinami liší!