

1 Naivný Bayesovský klasifikátor

1.1 Bude auto ukradnuté?

Máme nasledujúce empirické dáta o kradnutí áut.

Číslo	Farba	Typ	Pôvod	ukradnuté
1	červené	šport	tuzemské	áno
2	červené	šport	tuzemské	nie
3	červené	šport	tuzemské	áno
4	žlté	šport	tuzemské	nie
5	žlté	šport	dovoz	áno
6	žlté	kombi	dovoz	nie
7	žlté	kombi	dovoz	áno
8	žlté	kombi	tuzemské	nie
9	červené	kombi	dovoz	nie
10	červené	šport	dovoz	áno

Pomocou naivného Bayesovského klasifikátora určite, či hrozí, že červený kombík tuzemskej výroby bude ukradnutý.

1.2 Laplaceovská korekcia

Máme dáta o pacientoch po infarkte s nasledujúcimi položkami:

Vitamín pravidelne užívaný typ vitamínu.

Veľkosť rodiny veľkosť domácnosti, v ktorej pacient žil.

Cvičil či pravidelne cvičil.

Mal bypass či mal bypass.

Prežil či prežil aspoň 5 rokov od prvého infarktu.

Vitamín	Veľkosť rodiny	Cvičil	Mal bypass	Prežil
B complex	veľká	pravidelne	nie	—
B complex	veľká	pravidelne	áno	—
C	veľká	pravidelne	nie	+
D	stredná	pravidelne	nie	+
D	malá	málo	nie	+
D	malá	málo	áno	—
C	malá	málo	áno	+
B complex	stredná	pravidelne	nie	—
B complex	malá	málo	nie	+
D	stredná	málo	nie	+
B complex	stredná	málo	áno	+
C	stredná	pravidelne	áno	+
C	veľká	málo	nie	+
D	stredná	pravidelne	áno	—

Klasifikujte pacienta, ktorý

- berie vitamín D a má veľkú rodinu,
- berie vitamín D, má malú rodinu, pravidelne cvičí

Ako sa zmení výpočet naivného Bayesovského klasifikátora, ak použijeme Laplaceovskú korekciu. Bez tejto korekcie

$$P(E_k|H_t) = P(\text{vzor_z_triedy_t_vyhovuje_}E_k) = \frac{\text{Počet_vzorov_z_t_splňajúcich_}E_k}{\text{Počet_vzorov_z_t}}$$

s korekciou

$$P(E_k|H_t) = \frac{\text{Počet_vzorov_z_t_splňajúcich_}E_k + mp}{\text{Počet_vzorov_z_t} + m},$$

kde m je počet umelo pridaných vzorov a p je apriorná hodnota pravdepodobnosti $P(E_k|H_t)$. Veľmi často sa používa $m = T$ a $p = \frac{1}{T}$, kde T je počet tried (rôznych hodnôt pre t).