

Esercizio 1

Si consideri uno scenario di classificazione in cui si vuole predire se il capitale annuo supera i 50K e assumiamo che siano stati riportati i seguenti risultati su un'analisi effettuata secondo il sesso degli impiegati.

	FEMALE	MALE	TOTAL
>50K	100	230	330
<=50K	180	370	550
TOTAL	280	600	880

Per cui, supponendo di sapere che una persona guadagni più di 50K, qual è la probabilità che si tratti di una donna?

Soluzione

- La tabella indica delle donne considerate nel campione meno della metà, ovvero 100 su 280, hanno un guadagno annuo maggiore a 50K.
- Ma anche in generale poco meno della metà delle persone, ovvero 330 su 880, hanno un guadagno annuo maggiore a 50K.

Applicato il teorema di Bayes ai dati riportati in Tabella si deduce che:

$$P(F | HG) = \frac{P(HG | F)P(F)}{P(HG)} = \frac{0,357 * 0,318}{0,375} = 0,302 = 30,2\%$$

dove:

- $P(HG | F) = 100/280 = 0,357$
- $P(F) = 280/880 = 0,318$
- $P(HG) = 330/880 = 0,375$

Esercizio 5

Si consideri il seguente scenario di classificazione binaria in cui si vuole predire se il capitale annuo supera i 50K, indicato con Y/N

	Under 40 (f_1)	Graduated (f_2)	Single (f_3)	Capital
Training Data	1	1	1	N
	0	1	0	N
	1	1	1	N
	0	0	0	Y
	1	1	1	Y
	0	1	0	Y
Testing Case	1	0	0	?

Nella tabella sono inseriti sia i dati di training (Training Data) che il caso di test (Testing case) pari a $x = (1,0,0)$, calcolare le likelihood dato Y o N per ogni f_i con $i = 1,2,3$ applicando la tecnica di Laplace Smoothing.

Soluzione

- Sulla base della richiesta è necessario calcolare: $P(f_1 = 1|Y)$, $P(f_2 = 0|Y)$, $P(f_3 = 0|Y)$, $P(f_1 = 1|N)$, $P(f_2 = 0|N)$, $P(f_3 = 0|N)$
- $P(f_1 = 1|Y) = \frac{1+1}{3+2} = \frac{2}{5}$
- $P(f_2 = 0|Y) = \frac{1+1}{3+2} = \frac{2}{5}$
- $P(f_3 = 0|Y) = \frac{2+1}{3+2} = \frac{3}{5}$
- $P(f_1 = 1|N) = \frac{2+1}{3+2} = \frac{2}{5}$
- $P(f_2 = 0|N) = \frac{0+1}{3+2} = \frac{2}{5}$
- $P(f_3 = 0|N) = \frac{1+1}{3+2} = \frac{3}{5}$