

Teorema di Bayes $P(H|E) = \frac{P(E|H)P(H)}{P(E)}$

Supponiamo $C = \{c_1, c_2, \dots, c_m\}$

$$P(c_i|d) = \frac{P(d|c_i)P(c_i)}{P(d)}$$

Per classificare d mi interesso

$$\operatorname{argmax} P(c_i|d)$$

Perch' devo calcolare argmax posso riformulare il calcolo di $P(d)$

$P(d|c_i)$ e $P(c_i)$ li hanno stimate dal training set

$$TR = \{ \langle d_1, c_i \rangle \langle d_2, c_i \rangle \dots \langle d_k, c_i \rangle \}$$

$$P(c_i) = \frac{\# \text{ doc di classe } c_i \text{ nel TR}}{\# \text{ doc nel TR}}$$

$$P(d|c_i) = P(T_1 \wedge T_2 \wedge \dots \wedge T_k | c_i) \stackrel{\text{ASSUNZIONE NAIVE DI INDIPENDENZA}}{=} \prod_{j=1}^k P(T_j | c_i)$$

$$P(T_j | c_i) = \frac{\# \text{ volte in cui } T_j \text{ compare nei doc di classe } c_i \text{ nel TR} + 1}{\# \text{ totale dei termini nei doc di classe } c_i + |V|}$$

CORREZIONE DI
LAPLACE