Ragionamento Bayesiano
Esercizi

Classificatore Naïve Bayes: richiami (credits: P. Velardi-uniroma1)

- Si applica quando le ipotesi in H sono rappresentabili mediante una **congiunzione di valori di attributi** e la classificazione è scelta da un insieme finito Y. Le istanze **x** in X sono descritte da m-uple di valori (x_1, x_2, …, x_m) associati agli m attributi di **x**,

- si basa sull'assunzione semplificativa che I valori degli attributi siano **condizionalmente indipendenti**, assegnato un valore della funzione obiettivo.

Classificatore Naïve Bayes: richiami (2)

- Dato un nuovo esempio da classificare, si calcola:

$$c_{NB} = \arg\max_{c \in Y} P(c_j \mid x_1, x_2, \ldots, x_m) =$$

$$= \arg\max_{c \in Y} \frac{P(x_1, x_2, \ldots, x_m \mid c) \cdot P(c)}{P(x_1, x_2, \ldots, x_m)} =$$

$$= \arg\max_{c \in Y} P(c) \prod_j P(x_j \mid c)$$

Naïve Bayes: Esempio

- C = {allergia, raffreddore, in_salute} (**valori c(x))**
- att_1 = starnuti (sì, no) ; att_2 = tosse (sì, no) ; att_3 = febbre (sì, no) (attributi booleani)
- x = (1, 1, 0) ovvero (att_1, att_2, $\neg$ att_3) come lo classifico?

Dall'insieme D stimo le prob. a priori e condizionate es:

Prob	**in salute**	**raffreddore**	**allergia**
$P(c_i)$	0.9	0.05	0.05
$P(a_1 \mid c_i)$	0.027	1.0	1.0
$P(a_2 \mid c_i)$	0.027	0.5	0.5
$P(a_3 \mid c_i)$	0.027	0.5	0.5

Esempio (continua)

- 40 esempi, 36 classificati “in salute”, 2 “raffreddore”, 2 “allergia”
- Per stimare, ad esempio, P(a_1=1|in-salute), contare sui 36 esempi nei quali c(x)= “in-salute” quanti hanno att_1=1

  se 1 su 36, P(att_1=1|in-salute)=1/36=0,027

Analogamente avrò, ad es.:

- P(att_1=1|raffreddore)=2/2=1
- P(att_1=1|allergia)=2/2=1
- ecc.

Esempio (continua)

- Devo calcolare il massimo al variare di c di: $P(c_j)\prod_i P(a_i \mid c_j)$
- Quindi ad esempio per c=raffreddore

$$P(raffreddore)\left[P(att_1 = sì \mid raffr)P(att_2 = sì \mid raffr)P(att_3 = no \mid raffr)\right] =$$
$$0{,}05 \times \left[1 \times 0{,}5 \times 0{,}5\right] = 0{,}0125$$

- Analogamente, troverò:

$$P(in - salute)\left[P(att_1 = sì \mid sal)P(att_2 = sì \mid sal)P(att_3 = no \mid sal)\right] =$$
$$0{,}9 \times \left[0.027 \times 0{,}027 \times 0{,}027\right] = 0{,}000017$$
$$P(allergia)\left[P(att_1 = sì \mid all)P(att_2 = sì \mid all)P(att_3 = no \mid all)\right] =$$
$$0{,}05 \times \left[1 \times 0{,}5 \times 0{,}5\right] = 0{,}0125$$

Problemi con Naive Bayes

- Se D è piccolo, le stime sono inaffidabili (nell'esempio precedente alcune stime sono = 1!!!).
- Un valore raro a_k può non capitare mai in D e dunque:
  - $\forall c_j$: $P(a_k \mid c_j) = 0$.
- Analogamente, se ho un solo esempio di una classe c_j,
  - $\forall a_k$: $P(a_k \mid c_j) = 1$ o $P(a_k \mid c_j) = 0$.
- Se a_k capita in un test set T, il risultato è che
  - $\forall c_i$: $P(T \mid c_i) = 0$ and $\forall c_i$: $P(c_i \mid T) = 0$

  perché a_k non è mai apparso nel training set. Meglio adottare lo smoothing!!!

Smoothing: richiami

- Le probabilità $P(x_j|c)$ vengono stimate osservando le frequenze nei dati di addestramento D
- Se D include n_i esempi classificati c_i, e n_{ij} di questi n_i esempi contengono il valore x_j per l'attributo j, allora:

$$P(x_j \mid c_i) = \frac{n_{ij}}{n_i}$$

Smoothing: richiami (2)

- Per tener conto di eventi rari, si operano degli aggiustamenti sulle probabilità detti **smoothing**
- *Laplace smoothing con una M-stima* assume che ogni evento x_j abbia una probabilità a priori *p*, che si assume essere stata osservata in un campione virtuale di dimensione M > del campione reale

Smoothing: richiami (3)

$$P(x_j \mid c_i) = \frac{n_{ij} + Mp}{n_i + M}$$

- Nell'esempio precedente, si avrebbe:

$$P(x_1 = 0 \mid raff) = \frac{0 + M \times 0.5}{2 + M}$$

Smoothing: richiami (4)

- M è una costante che determina il peso dello smoothing
- In assenza di altre informazioni, si assume p=1/k dove k è il numero di valori dell’attributo j in esame.

Moneta bilanciata (credits: G. Manco)

- Lancio della moneta
  - Spazio: Ω = {Head, Tail}
  - Scenario: la moneta è bilanciata o sbilanciata al 60% in favore di Head
    - $h_1 \equiv$ bilanciata: P(Head) = 0.5 (più precisamente: P(Head | h_1))
    - $h_2 \equiv$ 60% bias: P(Head) = 0.6 (più precisamente: P(Head | h_2))
  - Obiettivo: decidere tra l'ipotesi di default (null) e l'alternativa.
- Distribuzione a-priori
  - $P(h_1) = 0.75$, $P(h_2) = 0.25$
  - Riflette le credenze iniziali su H
  - L'apprendimento è <u>revisione delle credenze</u>

Moneta bilanciata (continua)

- Evidenze
  - $d \equiv$ singolo lancio, viene Head
  - D: Cosa crediamo adesso?
  - R: Calcoliamo $P(d) = P(d|h_1) \cdot P(h_1) + P(d|h_2) \cdot P(h_2)$
- Inferenza Bayesiana: Calcolo di P(d=HEAD) $= P(d|h_1) \cdot P(h_1) + P(d|h_2) \cdot P(h_2) = 0.5 \bullet 0.75 + 0.6 \bullet 0.25 = 0.375 + 0.15 = 0.525$
- Questa è la probabilità dell’osservazione d=Head

Moneta bilanciata (continua)

- Apprendimento bayesiano
  - In base al teorema di Bayes
    - $P(h_1|d) = P(d|h_1)\ P(h_1) / P(d) = 0.375 / 0.525 = 0.714$
    - $P(h_2|d) = P(d|h_2)\ P(h_2) / P(d) = 0.15 / 0.525 = 0.286$
    - Le credenze sono state spostate verso h_1
    - MAP: crediamo ancora che la moneta sia bilanciata
  - Approccio ML (assumiamo priors identici)
    - Le credenze sono revisionate a partire da 0.5
    - C'è più sbilanciamento a favore di h_1

Moneta bilanciata (continua)

- Ulteriore evidenza: Sequenza D di 100 lanci con 70 heads e 30 tails
  - $P(D) = (0.5)^{70} \bullet (0.5)^{30} \bullet 0.75 + (0.6)^{70} \bullet (0.4)^{30} \bullet 0.25$
  - Ora $P(h_1|D) << P(h_2|D)$