Παράδειγμα 1

Σε κάποιο μάθημα εξετάστηκαν 250 άτομα, τα οποία τοποθετήθηκαν σε 3 αίθουσες A, B, Γ: 70 στην A, 90 στην B και τα υπόλοιπα στην Γ. Προβιβάσιμο βαθμό πήρε το 80% των γραπτών της αίθουσας A, το 88% της B και το 72% της Γ.

i) Αν επιλέξουμε τυχαία έναν εξεταζόμενο ποια είναι η πιθανότητα το γραπτό του να βαθμολογήθηκε κάτω από τη βάση;

ii) Αν πάρουμε ένα γραπτό που έχει βαθμολογηθεί πάνω από τη βάση, ποια είναι η πιθανότητα να προέρχεται από την αίθουσα A;

F = Το ενδεχομενο καποιος να πήρε μη προβιβασιμο βαθμο
A = Το ενδεχομενο να ειναι στην αιθουσα A
B = —η— B
Γ = —η— Γ

Απο την εκφώνηση γνωρίζουμε ότι

$P(A) = \frac{70}{250}$ $P(B) = \frac{90}{250}$ $P(\Gamma) = \frac{90}{250}$

$P(F|A) = 0.2$ $P(F|B) = 0.12$ $P(F|\Gamma) = 0.28$

i) Θέλουμε να υπολογίσουμε το $P(F)$

Απο τον τυπο της ολικής πιθανότητας

$$P(F) = P(A) \cdot P(F|A) + P(B) \cdot P(F|B) + P(\Gamma) \cdot P(F|\Gamma)$$

$$= \frac{70}{250} \cdot 0.2 + \frac{90}{250} \cdot 0.12 + \frac{90}{250} \cdot 0.28 = 0.2$$

ii) $P(A|\overline{F})$ $\overline{F}$ = Το ενδεχόμενο να μην προβλήθει με ψ βαθμό

Απο τον τυπο του Bayes $P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$ $B \neq \emptyset$

$$P(A\mid\overline{F}) = \frac{P(\overline{F}\mid A)\cdot P(A)}{P(\overline{F})}$$

$$= \frac{\left(1 - P(F\mid A)\right)\cdot P(A)}{1 - P(F)}$$

$$= \frac{0.8 \cdot \frac{70}{250}}{0.8} = \frac{70}{250}$$

Κανόνας του πολλαπλασιασμού

Όπως έχει ήδη αναφερθεί, τα ενδεχόμενα $A, B \subseteq \Omega$ ονομάζονται ανεξάρτητα αν και μόνο αν

$$P(A \cap B) = P(A)P(B).$$

Ο τύπος του Bayes δίνει άλλον έναν χαρακτηρισμό της ανεξαρτησίας δύο ενδεχομένων: Επειδή

$P(A)\cdot\cancel{P(B)}$

$$P(A|B) = P(A \cap B)/\cancel{P(B)} \qquad \text{και} \qquad P(B|A) = P(A \cap B)/P(A),$$

προκύπτει ότι τα ενδεχόμενα A, B είναι ανεξάρτητα αν και μόνο αν είναι $P(A|B) = P(A)$ (ή, ισοδύναμα, $P(B|A) = P(B)$).

Πρόταση.

Δύο ενδεχόμενα A, B είναι ανεξάρτητα αν και μόνο αν

$$P(A|B) = P(A) \qquad \text{ή} \qquad P(B|A) = P(B).$$

Κανόνας του πολλαπλασιασμού

Στην περίπτωση που τα ενδεχόμενα ενός δειγματικού χώρου δεν είναι κατ΄ ανάγκη ανεξάρτητα, ο παρακάτω τύπος είναι πολύ χρήσιμος.

Πρόταση.

Έστω $A_1, A_2, \ldots, A_n \subseteq \Omega$ με $P(A_1 \cap A_2 \cap \cdots \cap A_{n-1}) > 0$. Τότε

$$P\left(\bigcap_{i=1}^{n} A_i\right) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1})$$

Απόδειξη.

Ισχύει ότι

$$\begin{aligned}
&P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap A_2 \cap \cdots \cap A_{n-1}) \\
&= \cancel{P(A_1)} \frac{P(A_1 \cap A_2)}{\cancel{P(A_1)}} \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} \cdots \frac{P(A_1 \cap A_2 \cap \cdots \cap A_n)}{P(A_1 \cap A_2 \cap \cdots \cap A_{n-1})} \\
&= P(A_1 \cap A_2 \cap \cdots A_n).
\end{aligned}$$

□

Παράδειγμα - Έλεγχος ποιότητας

Ένας ελεγκτής ποιότητας εξετάζει μια παρτίδα με 100 τεμάχια ενός προϊόντος επιλέγοντας 5 τεμάχια (χωρίς επανατοποθέτηση). Αν κανένα από τα τεμάχια δεν είναι ελαττωματικό, τότε η παρτίδα γίνεται αποδεκτή, αλλιώς υποβάλλεται σε περαιτέρω ελέγχους. Να βρεθεί η πιθανότητα μια παρτίδα που περιέχει 5 ελαττωματικά τεμάχια να γίνει αποδεκτή.

Λύση

Έστω A_i το ενδεχόμενο το i-στό τεμάχιο που ελέγχεται να μην είναι ελαττωματικό, $i = 1, 2, 3, 4, 5$.
Μια παρτίδα γίνεται δεκτή αν πραγματοποιηθούν και τα 5 ενδεχόμενα, επομένως μας ενδιαφέρει η πιθανότητα $P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5)$.
Από τον κανόνα του πολλαπλασιασμού έχουμε ότι

$$P(A_1 \cap \cdots \cap A_5) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2)P(A_4|A_1 \cap A_2 \cap A_3) \cdot P(A_5|A_1 \cap A_2 \cap A_3 \cap A_4).$$

Εστω οτι για παρτιδα περιεχει 14 ελαττωματικα κομματια

Ποια ειναι η πιθανοτητα να περασει τον ελεγχο;

A_i = Το ενδεχομενο το i-οστο αντικειμενο που ελεγχουμε να μην ειναι ελαττωματικο

Η ζητουμενη πιθανοτητα ειναι (απο τον τυπο του γινομενου)

$$P(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5) =$$

$$P(A_1) \cdot P(A_2 | A_1) \cdot P(A_3 | A_1 \cap A_2) \cdot P(A_4 | A_1 \cap A_2 \cap A_3) \cdot P(A_5 | A_1 \cap A_2 \cap A_3 \cap A_4)$$

$$= \frac{86}{100} \cdot \frac{85}{99} \cdot \frac{84}{98} \cdot \frac{83}{97} \cdot \frac{82}{96} = 0.46$$

Ανεξαρτησία υπό συνθήκη

Επειδή η δεσμευμένη πιθανότητα είναι συνάρτηση πιθανότητας μπορούμε να μιλάμε και για ανεξαρτησία ενδεχομένων υπό συνθήκη (conditional independence).

Ορισμός - Ανεξαρτησία υπό συνθήκη.

Έστω $\emptyset \neq C \subseteq \Omega$. Τα $A, B \subseteq \Omega$ ονομάζονται **υπό συνθήκη** ή **υπό δέσμευση ανεξάρτητα δεδομένου του** C αν και μόνο αν

$$P(A \cap B | C) = P(A|C)P(B|C).$$

Παρατηρήστε ότι γενικά

$$\begin{aligned} P(A \cap B|C) &= P(A \cap B \cap C)/P(C) = P(C)P(B|C)P(A|B \cap C)/P(C) \\ &= P(B|C)P(A|B \cap C). \end{aligned}$$

Για να είναι τα A, B υπο συνθήκη ανεξάρτητα δεδομένου του C πρέπει

$$P(A|B \cap C) = P(A|C)$$

δηλαδή αν είναι γνωστό ότι το C έχει συμβεί, η επιπλέον πληροφορία ότι το B έχει επίσης συμβεί να μην αλλάζει την πιθανότητα του A.

Ανεξαρτησία υπό συνθήκη

Παράδειγμα

Για το παράδειγμα 71, να βρεθεί η πιθανότητα κάποιος άνθρωπος να πάσχει από αυτή την πάθηση, δεδομένου ότι υποβλήθηκε σε δύο διαδοχικά τεστ, τα οποία βγήκαν θετικά.

Λύση

Υπενθυμίζεται ότι το ποσοστό του πληθυσμού που έχει την πάθηση είναι $x = 0.002$ και ότι το τεστ δίνει σωστό αποτέλεσμα με πιθανότητα $p = P(\Theta|A) = P(\overline{\Theta}|\overline{A}) = 0.95$. Συμβολίζουμε με Θ_1 και Θ_2 τα ενδεχόμενα να βγουν θετικά το πρώτο και το δεύτερο τεστ αντίστοιχα. Τα ενδεχόμενα Θ_1 και Θ_2 δεν είναι ανεξάρτητα, αλλά είναι ανεξάρτητα δεδομένου του A, δηλαδή $P(\Theta_1 \cap \Theta_2|A) = P(\Theta_1|A)P(\Theta_2|A) = p^2$. Με απλά λόγια, αν κάποιος που είναι ασθενής υποβληθεί σε διαδοχικά τεστ, το κάθε ένα από αυτά θα βγαίνει θετικό με πιθανότητα $p = 0.95$ και ανεξάρτητα από τα υπόλοιπα.

A = Το ενδεχόμενο κάποιος να είναι ασθενής

$$P(A) = \frac{2}{1000}$$

$$P(\Theta \mid A) = P(\bar{\Theta} \mid \bar{A}) = 0.95$$

$$P(\bar{\Theta} \mid A) = P(\Theta \mid \bar{A}) = 0.05$$

$$P(A \mid \Theta_1 \cap \Theta_2) = ?$$

Από τον τύπο του Bayes

$$P(A \mid \Theta_1 \cap \Theta_2) = \frac{P(\Theta_1 \cap \Theta_2 \mid A)\, P(A)}{P(\Theta_1 \cap \Theta_2)}$$

Τα Θ_1, Θ_2 δεν είναι ανεξάρτητα από μόνα τους
αλλά είναι ανεξάρτητα δεδομένου του A (ή του $\bar{A}$

$$P(\Theta_1 \cap \Theta_2 \mid A) = P(\Theta_1 \mid A) \cdot P(\Theta_2 \mid A)$$

Άρα, έχουμε

$$P(\Theta_1 \cap \Theta_2 | A) \cdot P(A) = P(\Theta_1|A) \cdot P(\Theta_2|A) \cdot P(A)$$

$$= 0.95 \cdot 0.95 \cdot \frac{2}{1000}$$

Επιπλεον, απο τον τυπο της ολικης πιθανοτητας

$$P(\Theta_1 \cap \Theta_2) = P(A) \cdot P(\Theta_1 \cap \Theta_2 | A) + P(\bar{A}) \cdot P(\Theta_1 \cap \Theta_2 | \bar{A})$$

$$= P(A) \cdot P(\Theta_1|A) P(\Theta_2|A) + P(\bar{A}) \cdot P(\Theta_1|\bar{A}) \cdot P(\Theta_2|\bar{A})$$

$$= \frac{2}{1000} \cdot 0.95 \cdot 0.95 + \frac{998}{1000} \cdot 0.05 \cdot 0.05$$

Άρα

$$P(A|\Theta_1 \cap \Theta_2) = \frac{0.95^2 \cdot \frac{2}{1000}}{0.95^2 \cdot \frac{2}{1000} + 0.05^2 \cdot \frac{998}{1000}} = 0.42$$

Naive bayes classification

2. Για μια ομάδα ατόμων έχουμε συλλέξει τα παρακάτω δεδομένα

age	income	student	credit rating	buys computer
≤ 30	high	no	fair	no
≤ 30	high	no	excellent	no
31..40	high	no	fair	yes
> 40	medium	no	fair	yes
> 40	low	yes	fair	yes
> 40	low	yes	excellent	no
31..40	low	yes	excellent	yes
≤ 30	medium	no	fair	no
≤ 30	low	yes	fair	yes
> 40	medium	yes	fair	yes
≤ 30	medium	yes	excellent	yes
31..40	medium	no	excellent	yes
31..40	high	yes	fair	yes
> 40	medium	no	excellent	no

#yes = 9

#No = 5

Με βάση τα παραπάνω δεδομένα που θα κατηγοριοποιούσαμε ένα άτομο με τα παρακάτω χαρακτηριστικά

age	income	student	credit rating	buys computer
> 40	high	no	excellent	?

A_1 = Το ενδεχομενο η ηλικια να ειναι ≥ 40

I_1 = Το ενδεχομενο το εισοδημα να ειναι high
S_1 = Το ενδεχομενο καποιος να μη σπουδαζει
C_1 = Το ενδεχομενο καποιος να εχει excellent credit rating
B = Το ενδεχομενο να αγοράσει το προϊόν

$$P(B \mid A_1 \cap I_1 \cap S_1 \cap C_1) = 1$$
$$P(\bar{B} \mid A_1 \cap I_1 \cap S_1 \cap C_1) = 2$$

Ιδεα: Θα βρουμε ποιο ειναι πιθανο και θα θεωρησουμε οτι θα συμβει αυτο που ειναι πιθανοτερ

$$1 = \frac{P(A_1 \cap I_1 \cap S_1 \cap C_1 \mid B)\, P(B)}{P(A_1 \cap I_1 \cap S_1 \cap C_1)}$$

$$2 = \frac{P(A_1 \cap I_1 \cap S_1 \cap C_1 \mid \bar{B})\, P(\bar{B}}{P(A_1 \cap I_1 \cap S_1 \cap C_1)}$$

Τα δυο κλασματα εχουν κοινο παρονομαστη οποτε για να αποφασισω αρκει να συγκρινω αριθμητες

Υποθετουμε* οτι τα A_1, I_1, S_1, C_1 ειναι ανεξαρτητα δεδομενου του B (και του $\bar{B}$)

Οποτε για το ① εχουμε ως αριθμητη

$$P(A_1|B)\cdot P(I_1|B)\cdot P(S_1|B)\cdot P(C_1|B)\cdot P(B)$$

$$\frac{3}{9}\cdot\frac{2}{9}\cdot\frac{3}{9}\cdot\frac{3}{9}\cdot\frac{9}{14} = \frac{1}{189} = 0{,}005$$

(δεν ειναι πιθανοτητα)

Για το ② εχουμε ως αριθμητη

$$P(A_1|\bar{B})\cdot P(I_1|\bar{B})\cdot P(S_1|\bar{B})\cdot P(C_1|\bar{B})\cdot P(\bar{B})$$

$$\frac{2}{5}\cdot\frac{2}{5}\cdot\frac{4}{5}\cdot\frac{3}{5}\cdot\frac{5}{14} = \frac{24}{875} = 0{,}027$$

Επειδη ② > ① αρα ειναι πιο πιθανο ου θα ισχυει το $\bar{B}$.