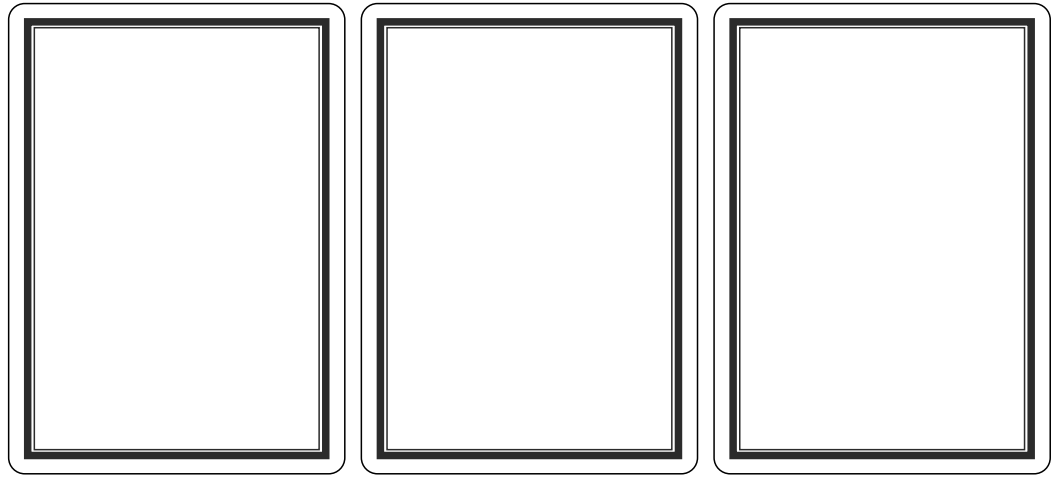
Μάθηση βασισμένη σε πιθανότητες

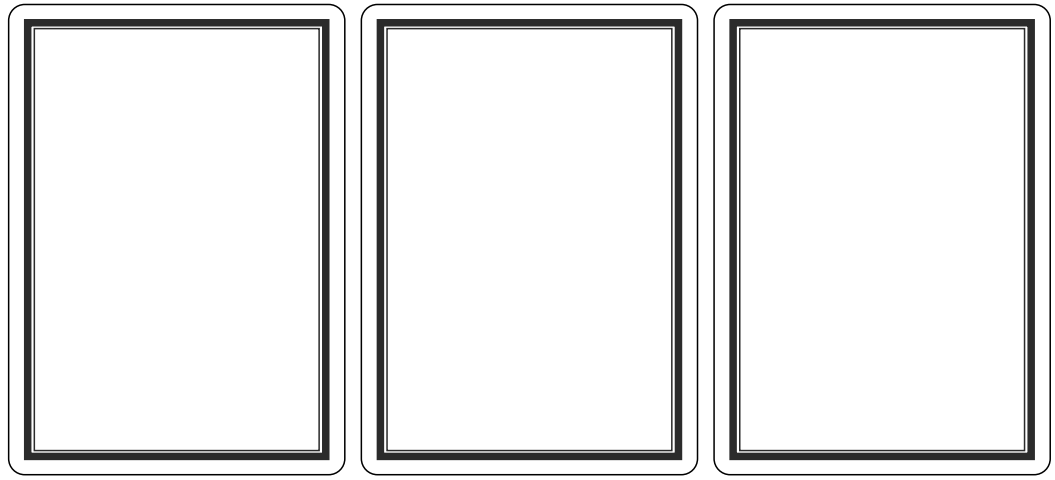
Κεντρική Ιδέα

(α΄)

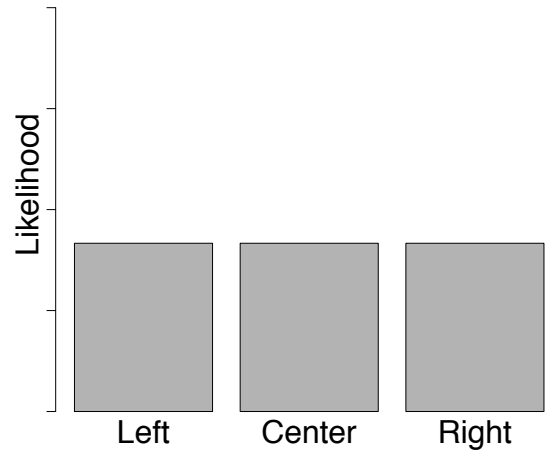


(β΄)

Σχήμα: Ένα παιχνίδι *βρες τη ντάμα*

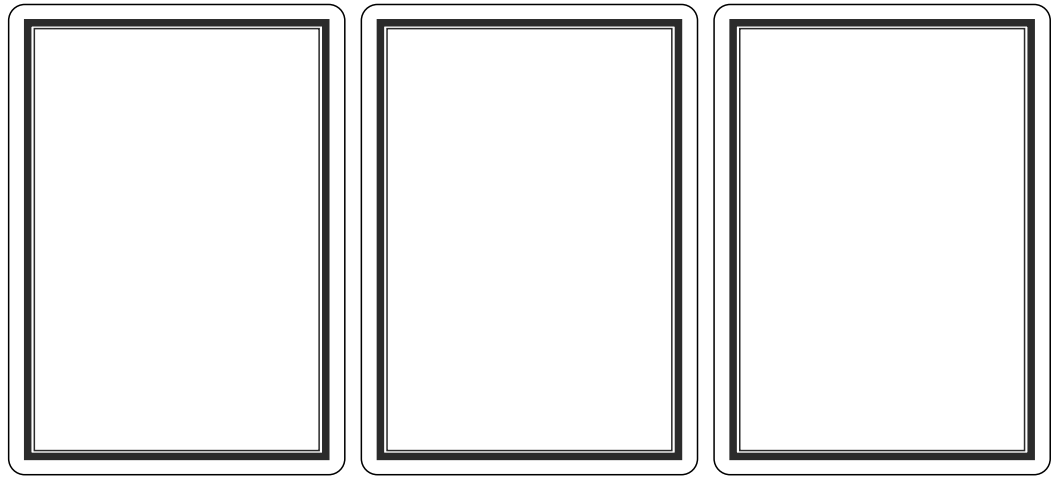


(α΄)

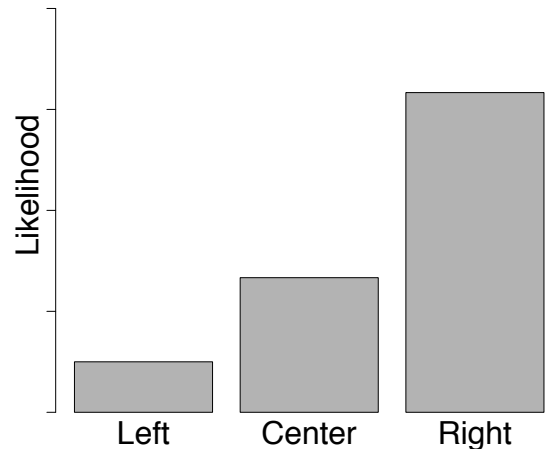


(β΄)

Σχήμα: Ένα παιχνίδι *βρες τη ντάμα*: (a) τα χαρτιά μοιρασμένα κλειστά πάνω στο τραπέζι· και (b) οι αρχικές πιθανότητες η ντάμα να καταλήξει σε κάθε θέση.

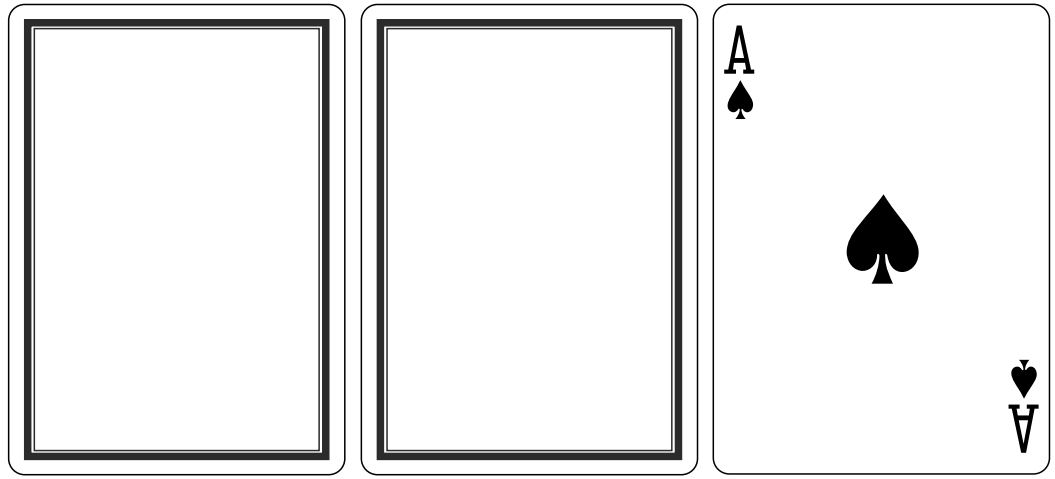


(α΄)

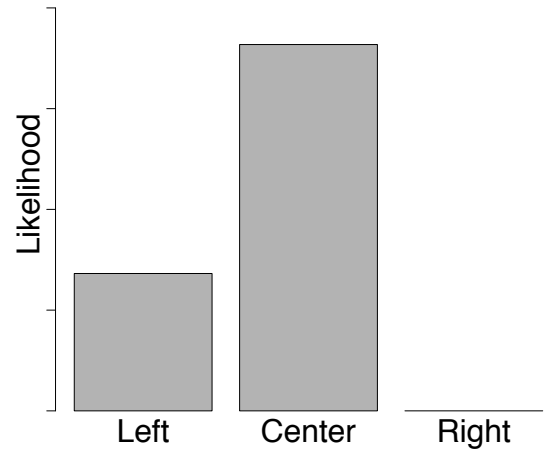


(β΄)

Σχήμα: Ένα παιχνίδι *βρες τη ντάμα*: (a) τα χαρτιά μοιρασμένα κλειστά πάνω στο τραπέζι· και (b) ένα αναθεωρημένο σύνολο πιθανοτήτων για τη θέση της ντάμας με βάση τα δεδομένα που συλλέχθηκαν.

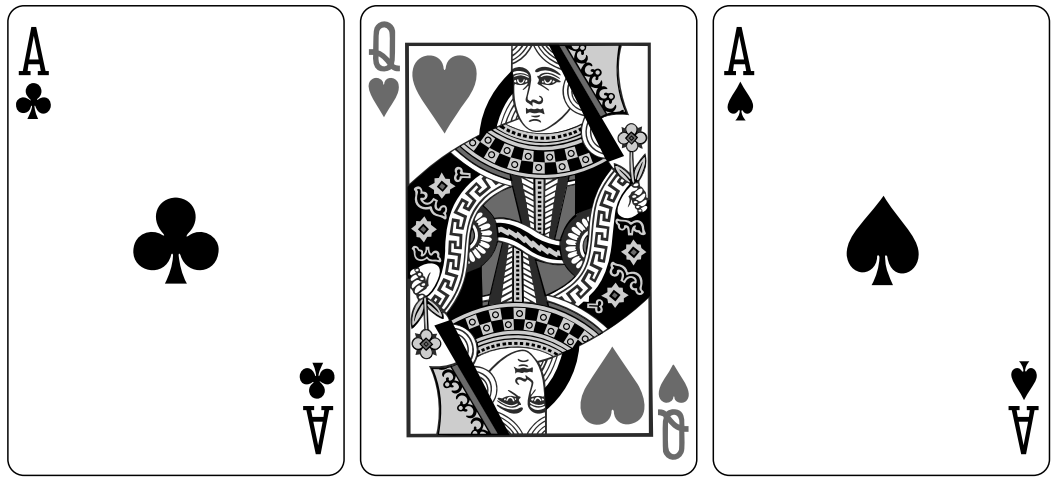


(α΄)



(β΄)

Σχήμα: Ένα παιχνίδι *βρες τη ντάμα*: (a) το σύνολο των χαρτιών αφού ο αέρας αναποδογυρίζει εκείνο στα δεξιά· (b) οι αναθεωρημένες πιθανότητες για τη θέση της ντάμας με βάση αυτό το νέο στοιχείο.



Σχήμα: Ένα παιχνίδι *βρες τη ντάμα*: οι τελικές θέσεις των χαρτιών στο παιχνίδι.

Κεντρική Ιδέα

- Μπορούμε να χρησιμοποιούμε εκτιμήσεις πιθανοτήτων για να προσδιορίσουμε ποια είναι η πιο πιθανή πρόβλεψη που πρέπει να γίνει.
- Ακόμη πιο σημαντικό, αναθεωρούμε αυτές τις προβλέψεις με βάση τα δεδομένα που συλλέγουμε και κάθε φορά που γίνεται διαθέσιμο νέο αποδεικτικό στοιχείο.

Θεμελιώδεις Έννοιες

Πίνακας: Ένα απλό σύνολο δεδομένων για τη διάγνωση Meningitis, με περιγραφικά χαρακτηριστικά που περιγράφουν την παρουσία ή απουσία τριών κοινών συμπτωμάτων της νόσου: Headache, Fever και Vomiting.

ID	Headache	Fever	Vomiting	Meningitis
1	true	true	false	false
2	false	true	false	false
3	true	false	true	false
4	true	false	true	false
5	false	true	false	true
6	true	false	true	false
7	true	false	true	false
8	true	false	true	true
9	false	true	false	false
10	true	false	true	true

Πιθανοτική ερμηνεία ενός συνόλου δεδομένων

- Κάθε χαρακτηριστικό → **τυχαία μεταβλητή**
- Όλοι οι πιθανοί συνδυασμοί τιμών → **δειγματικός χώρος**
- Κάθε γραμμή του συνόλου δεδομένων → **πείραμα**
- Κάθε συγκεκριμένη αντιστοίχιση τιμών στα περιγραφικά χαρακτηριστικά → **συμβάν**

Παράδειγμα

Κάθε γραμμή στον Πίνακα 6.1 αποτελεί ένα πείραμα και οι τιμές των περιγραφικών χαρακτηριστικών της γραμμής ορίζουν ένα διακριτό συμβάν.

Συνάρτηση πιθανότητας

- Μια **συνάρτηση πιθανότητας** (probability function) $P()$ επιστρέφει την πιθανότητα ενός συμβάντος.
- Μια **συνάρτηση πιθανότητας**, $P()$, επιστρέφει την πιθανότητα ένα χαρακτηριστικό να λάβει μια συγκεκριμένη τιμή.
- Για παράδειγμα, η $P(\text{FEVER} = \textit{true})$ επιστρέφει την πιθανότητα το χαρακτηριστικό FEVER να πάρει την τιμή *'true'*.
- Η πιθανότητα αυτή, η οποία είναι 0.4, μπορεί να μετρηθεί απευθείας από το σύνολο δεδομένων.
- Οι συναρτήσεις πιθανότητας για **κατηγορικά χαρακτηριστικά** ονομάζονται **συναρτήσεις μάζας πιθανότητας** (probability mass functions).
- Οι συναρτήσεις πιθανότητας για **συνεχή χαρακτηριστικά** ονομάζονται **συναρτήσεις πυκνότητας πιθανότητας** (probability density functions).

- Μια **μικτή πιθανότητα** αναφέρεται στην πιθανότητα ανάθεσης συγκεκριμένων τιμών σε πολλά διαφορετικά χαρακτηριστικά, για παράδειγμα P(Meningitis = true,Headache = true) = 0,2.
- Μια **δεσμευμένη πιθανότητα** αναφέρεται στην πιθανότητα ένα χαρακτηριστικό να λάβει μια συγκεκριμένη τιμή, δεδομένου ότι γνωρίζουμε ήδη την τιμή ενός άλλου χαρακτηριστικού, όπως για παράδειγμα P(Meningitis = true | Headache = true) = 0,2857.
- Μια **κατανομή πιθανότητας** είναι μια δομή δεδομένων που περιγράφει την πιθανότητα κάθε δυνατής τιμής που μπορεί να λάβει ένα χαρακτηριστικό. Το άθροισμα μιας κατανομής πιθανότητας πρέπει να ισούται με 1.0.

- Μια **κατανομή μικτής πιθανότητας** είναι μια κατανομή πιθανότητας σε περισσότερες από μία αναθέσεις χαρακτηριστικών και γράφεται ως πολυδιάστατος πίνακας, στον οποίο κάθε κελί δίνει την πιθανότητα ενός συγκεκριμένου συνδυασμού τιμών χαρακτηριστικών.
- Το άθροισμα όλων των κελιών μιας κατανομής μικτής πιθανότητας πρέπει να ισούται με 1.0.

$$\mathbf{P}(H, F, V, M) = \begin{bmatrix} P(h, f, v, m), & P(\neg h, f, v, m) \\ P(h, f, v, \neg m), & P(\neg h, f, v, \neg m) \\ P(h, f, \neg v, m), & P(\neg h, f, \neg v, m) \\ P(h, f, \neg v, \neg m), & P(\neg h, f, \neg v, \neg m) \\ P(h, \neg f, v, m), & P(\neg h, \neg f, v, m) \\ P(h, \neg f, v, \neg m), & P(\neg h, \neg f, v, \neg m) \\ P(h, \neg f, \neg v, m), & P(\neg h, \neg f, \neg v, m) \\ P(h, \neg f, \neg v, \neg m), & P(\neg h, \neg f, \neg v, \neg m) \end{bmatrix}$$

- Δοθείσης μιας κατανομής μικτής πιθανότητας, μπορούμε να υπολογίσουμε την πιθανότητα οποιουδήποτε γεγονότος στο πεδίο που καλύπτει, αθροίζοντας τα κελιά της κατανομής στα οποία το γεγονός αυτό είναι αληθές.
- Για παράδειγμα, για να υπολογίσουμε την πιθανότητα του P(h) στο πεδίο που καθορίζεται από την κατανομή μικτών πιθανοτήτων P(H,F,V,M), απλώς αθροίζουμε τις τιμές στα κελιά που περιέχουν το h (τα κελιά στην πρώτη στήλη).
- Ο υπολογισμός πιθανοτήτων με αυτόν τον τρόπο ονομάζεται **άθροιση ως προς** (*summing out*).
- Μπορούμε επίσης να χρησιμοποιήσουμε την άθροιση για να υπολογίσουμε τις δεσμευμένες πιθανότητες από μια κατανομή μικτών πιθανοτήτων. Για να υπολογίσουμε την πιθανότητα P(h | f) από το P(H,F,V,M), αθροίζουμε τις τιμές σε όλα τα κελιά όπου ισχύουν τα h και f (τα τέσσερα πρώτα κελιά στην πρώτη στήλη).
- Είμαστε έτοιμοι να ασχοληθούμε με το Θεώρημα του Bayes!

Το Θεώρημα του Bayes

$$P(X|Y) = \frac{P(Y|X)P(X)}{P(Y)}$$

Example

Μετά από έναν ετήσιο έλεγχο, ένας γιατρός ενημερώνει τον ασθενή του ότι έχει και κακά νέα και καλά νέα. Τα κακά νέα είναι ότι ο ασθενής βρέθηκε θετικός σε μια σοβαρή ασθένεια και ότι το τεστ που χρησιμοποίησε ο γιατρός είναι ακριβές κατά 99% (δηλαδή, η πιθανότητα θετικού αποτελέσματος όταν ένας ασθενής έχει την ασθένεια είναι 0.99, όπως και η πιθανότητα αρνητικού αποτελέσματος όταν ο ασθενής δεν έχει την ασθένεια). Τα καλά νέα, ωστόσο, είναι ότι η ασθένεια είναι εξαιρετικά σπάνια και προσβάλλει μόνο 1 στους 10,000 ανθρώπους.

- Ποια είναι η πραγματική πιθανότητα ο ασθενής να έχει την ασθένεια;
- Γιατί η σπανιότητα της ασθένειας είναι καλό νέο, δεδομένου ότι ο ασθενής βρέθηκε θετικός;

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

$$\begin{aligned} P(t) &= P(t|d)P(d) + P(t|\neg d)P(\neg d) \\ &= (0.99 \times 0.0001) + (0.01 \times 0.9999) = 0.0101 \end{aligned}$$

$$\begin{aligned} P(d|t) &= \frac{0.99 \times 0.0001}{0.0101} \\ &= 0.0098 \end{aligned}$$

Παράγωγη του θεωρήματος του Bayes

$$P(Y|X)\,P(X) = P(X|Y)\,P(Y)$$

$$\frac{P(X|Y)\,P(Y)}{P(Y)} = \frac{P(Y|X)\,P(X)}{P(Y)}$$

$$\frac{P(X|Y)\,\cancel{P(Y)}}{\cancel{P(Y)}} = \frac{P(Y|X)\,P(X)}{P(Y)}$$

$$\Rightarrow P(X|Y) = \frac{P(Y|X)\,P(X)}{P(Y)}$$

- Ο διαιρέτης είναι η εκ των προτέρων πιθανότητα του αποδεικτικού στοιχείου.
- Αυτή η διαίρεση λειτουργεί ως σταθερά κανονικοποίησης.

$$\begin{aligned} 0 \leq P(X|Y) &\leq 1 \\ \sum_i P(X_i|Y) &= 1.0 \end{aligned}$$

- Μπορούμε να υπολογίσουμε αυτόν τον διαιρέτη απευθείας από το σύνολο δεδομένων.

$$P(Y) = \frac{|\{\text{γραμμές όπου ισχύει το Y}\}|}{|\{\text{γραμμές στο σύνολο δεδομένων}\}|}$$

- Ή μπορούμε να χρησιμοποιήσουμε το **Θεώρημα της Ολικής Πιθανότητας** για να υπολογίσουμε αυτόν τον διαιρέτη.

$$P(Y) = \sum_i P(Y|X_i)P(X_i) \tag{1}$$

Για να κάνουμε προβλέψεις κατά Bayes, δημιουργούμε την πιθανότητα του συμβάντος όπου ένα χαρακτηριστικό-στόχος, το t, θα πάρει ένα συγκεκριμένο επίπεδο, το l, δοθείσης της ανάθεσης τιμών σε ένα σύνολο περιγραφικών χαρακτηριστικών q, από ένα στιγμιότυπο ερωτήματος.

Γενικευμένο Θεώρημα του Bayes

$$P(t = l|\mathbf{q}[1], \ldots, \mathbf{q}[m]) = \frac{P(\mathbf{q}[1], \ldots, \mathbf{q}[m]|t = l)P(t = l)}{P(\mathbf{q}[1], \ldots, \mathbf{q}[m])}$$

Για να υπολογίσουμε μια πιθανότητα χρησιμοποιώντας το Γενικευμένο Θεώρημα Bayes, πρέπει να υπολογίσουμε τρεις πιθανότητες:

- την P(t = l), δηλαδή την προηγούμενη πιθανότητα (prior probability) του χαρακτηριστικού στόχου t που έχει το επίπεδο l
- την P(q[1],...,q[m]), δηλαδή τη μικτή πιθανότητα των περιγραφικών χαρακτηριστικών ενός στιγμιοτύπου ερωτήματος που παίρνει ένα συγκεκριμένο σύνολο τιμών
- την P(q[1],...,q[m] | t = l), δηλαδή τη δεσμευμένη πιθανότητα των περιγραφικών χαρακτηριστικών ενός στιγμιοτύπου ερωτήματος που παίρνει ένα συγκεκριμένο επίπεδο τιμών, δεδομένου του χαρακτηριστικού-στόχου που έχει το επίπεδο l

Κανόνας Αλυσίδας

$$P(\mathbf{q}[1], \ldots, \mathbf{q}[m]) = P(\mathbf{q}[1]) \times P(\mathbf{q}[2]|\mathbf{q}[1]) \times \cdots \times P(\mathbf{q}[m]|\mathbf{q}[m-1], \ldots, \mathbf{q}[2], \mathbf{q}[1])$$

- Για να εφαρμόσουμε τον κανόνα αλυσίδας σε μια δεσμευμένη πιθανότητα, απλώς προσθέτουμε τον όρο δέσμευσης σε κάθε όρο της έκφρασης:

$$P(\mathbf{q}[1], \ldots, \mathbf{q}[m]|t = l) = P(\mathbf{q}[1]|t = l) \times P(\mathbf{q}[2]|\mathbf{q}[1], t = l) \times \ldots \cdots \times P(\mathbf{q}[m]|\mathbf{q}[m-1], \ldots, \mathbf{q}[3], \mathbf{q}[2], \mathbf{q}[1], t = l)$$

ID	Headache	Fever	Vomiting	Meningitis
1	true	true	false	false
2	false	true	false	false
3	true	false	true	false
4	true	false	true	false
5	false	true	false	true
6	true	false	true	false
7	true	false	true	false
8	true	false	true	true
9	false	true	false	false
10	true	false	true	true

Headache	Fever	Vomiting	Meningitis
true	false	true	?

$$P(M|h, \neg f, v) = ?$$

- Με όρους του θεωρήματος του Bayes, αυτό το πρόβλημα μπορεί να διατυπωθεί ως εξής:

$$P(M|h, \neg f, v) = \frac{P(h, \neg f, v|M) \times P(M)}{P(h, \neg f, v)}$$

- Υπάρχουν δύο τιμές στο πεδίο του χαρακτηριστικού MENINGITIS, *'true'* και *'false'*, οπότε πρέπει να κάνουμε αυτόν τον υπολογισμό δύο φορές.

- Θα κάνουμε πρώτα τον υπολογισμό για το m
- Για να πραγματοποιήσουμε αυτόν τον υπολογισμό πρέπει να γνωρίζουμε τις εξής πιθανότητες: $P(m)$, $P(h, \neg f, v)$ και $P(h, \neg f, v \mid m)$.

ID	Headache	Fever	Vomiting	Meningitis
1	true	true	false	false
2	false	true	false	false
3	true	false	true	false
4	true	false	true	false
5	false	true	false	true
6	true	false	true	false
7	true	false	true	false
8	true	false	true	true
9	false	true	false	false
10	true	false	true	true

- Μπορούμε να υπολογίσουμε τις απαιτούμενες πιθανότητες απευθείας από τα δεδομένα. Για παράδειγμα, μπορούμε να υπολογίσουμε τα $P(m)$ και $P(h, \neg f, v)$ ως εξής:

$$P(m) = \frac{|\{\mathbf{d}_5, \mathbf{d}_8, \mathbf{d}_{10}\}|}{|\{\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3, \mathbf{d}_4, \mathbf{d}_5, \mathbf{d}_6, \mathbf{d}_7, \mathbf{d}_8, \mathbf{d}_9, \mathbf{d}_{10}\}|} = \frac{3}{10} = 0.3$$

$$P(h, \neg f, v) = \frac{|\{\mathbf{d}_3, \mathbf{d}_4, \mathbf{d}_6, \mathbf{d}_7, \mathbf{d}_8, \mathbf{d}_{10}\}|}{|\{\mathbf{d}_1, \mathbf{d}_2, \mathbf{d}_3, \mathbf{d}_4, \mathbf{d}_5, \mathbf{d}_6, \mathbf{d}_7, \mathbf{d}_8, \mathbf{d}_9, \mathbf{d}_{10}\}|} = \frac{6}{10} = 0.6$$

- Ωστόσο, ως άσκηση θα χρησιμοποιήσουμε τον κανόνα αλυσίδας για να υπολογίσουμε:

$$P(h, \neg f, v \mid m) = ?$$

ID	Headache	Fever	Vomiting	Meningitis
1	true	true	false	false
2	false	true	false	false
3	true	false	true	false
4	true	false	true	false
5	false	true	false	true
6	true	false	true	false
7	true	false	true	false
8	true	false	true	true
9	false	true	false	false
10	true	false	true	true

- Χρησιμοποιώντας τον κανόνα αλυσίδας υπολογίστε:

$$
\begin{aligned}
P(h, \neg f, v \mid m) &= P(h \mid m) \times P(\neg f \mid h, m) \times P(v \mid \neg f, h, m) \\
&= \frac{|\{\mathbf{d}_8, \mathbf{d}_{10}\}|}{|\{\mathbf{d}_5, \mathbf{d}_8, \mathbf{d}_{10}\}|} \times \frac{|\{\mathbf{d}_8, \mathbf{d}_{10}\}|}{|\{\mathbf{d}_8, \mathbf{d}_{10}\}|} \times \frac{|\{\mathbf{d}_8, \mathbf{d}_{10}\}|}{|\{\mathbf{d}_8, \mathbf{d}_{10}\}|} \\
&= \frac{2}{3} \times \frac{2}{2} \times \frac{2}{2} = 0.6666
\end{aligned}
$$

- Άρα ο υπολογισμός του $P(m|h, \neg f, v)$ είναι:

$$P(m|h, \neg f, v) = \frac{\left(\begin{array}{l} P(h|m) \times P(\neg f|h, m) \\ \quad \times P(v|\neg f, h, m) \times P(m) \end{array}\right)}{P(h, \neg f, v)}$$
$$= \frac{0.6666 \times 0.3}{0.6} = 0.3333$$

- Ο αντίστοιχος υπολογισμός για το $P(\neg m|h, \neg f, v)$ είναι:

$$P(\neg m \mid h, \neg f, v) = \frac{P(h, \neg f, v \mid \neg m) \times P(\neg m)}{P(h, \neg f, v)}$$

$$= \frac{\begin{pmatrix} P(h|\neg m) \times P(\neg f \mid h, \neg m) \\ \quad \times P(v|\neg f, h, \neg m) \times P(\neg m) \end{pmatrix}}{P(h, \neg f, v)}$$

$$= \frac{0.7143 \times 0.8 \times 1.0 \times 0.7}{0.6} = 0.6667$$

$$P(m|h, \neg f, v) = 0.3333$$

$$P(\neg m|h, \neg f, v) = 0.6667$$

- Αυτοί οι υπολογισμοί μάς λένε ότι είναι δύο φορές πιο πιθανό ο ασθενής να μην έχει μηνιγγίτιδα απ' ό,τι να έχει, παρόλο που ο ασθενής υποφέρει από πονοκέφαλο και κάνει εμετό!

Το Παράδοξο του Ψευδώς Θετικού

- Το λάθος του να ξεχνάμε να λάβουμε υπόψη την εκ των προτέρων πιθανότητα οδηγεί στο **παράδοξο του ψευδώς θετικού**, το οποίο δηλώνει ότι, για να γίνουν προβλέψεις σχετικά με ένα σπάνιο γεγονός, το μοντέλο πρέπει να είναι τόσο ακριβές όσο σπάνια είναι η εκ των προτέρων πιθανότητα του γεγονότος, αλλιώς υπάρχει σημαντική πιθανότητα **ψευδώς θετικών** προβλέψεων (δηλαδή, πρόβλεψη ότι το γεγονός συμβαίνει ενώ στην πραγματικότητα δεν συμβαίνει).

- Ο ακριβής υπολογισμός των πιθανοτήτων για κάθε ένα από τα πιθανά επίπεδα-στόχους είναι συχνά πολύ χρήσιμος για έναν άνθρωπο που λαμβάνει αποφάσεις, όπως για παράδειγμα έναν γιατρό.
- Ωστόσο, αν προσπαθούμε να κατασκευάσουμε ένα προγνωστικό μοντέλο που αναθέτει αυτόματα ένα επίπεδο-στόχο σε ένα στιγμιότυπο ερωτήματος, τότε πρέπει να αποφασίσουμε πώς το μοντέλο θα κάνει μια πρόβλεψη με βάση τις υπολογισμένες πιθανότητες.
- Ο προφανής τρόπος για να το κάνουμε αυτό είναι να Cητήσουμε από το μοντέλο να επιστρέψει το επίπεδο-στόχο που έχει την υψηλότερη εκ των υστέρων πιθανότητα, δοθείσης της κατάστασης των περιγραφικών χαρακτηριστικών στο ερώτημα.
- Ένα μοντέλο πρόβλεψης που λειτουργεί με αυτόν τον τρόπο κάνει μια μέγιστη εκ των υστέρων (maximum a posteriori (MAP)) πρόβλεψη.

Μοντέλο Μπεϋζιανής Πρόβλεψης MAP

$$\begin{aligned}\mathbb{M}_{MAP}(\mathbf{q}) &= \underset{l \in levels(t)}{\operatorname{argmax}} P(t = l \mid \mathbf{q}[1], \ldots, \mathbf{q}[m]) \\ &= \underset{l \in levels(t)}{\operatorname{argmax}} \frac{P(\mathbf{q}[1], \ldots, \mathbf{q}[m] \mid t = l) \times P(t = l)}{P(\mathbf{q}[1], \ldots, \mathbf{q}[m])}\end{aligned}$$

Μοντέλο Μπεϋζιανής Πρόβλεψης MAP (χωρίς κανονικοποίηση)

$$\mathbb{M}_{MAP}(\mathbf{q}) = \underset{l \in levels(t)}{\operatorname{argmax}} P(\mathbf{q}[1], \ldots, \mathbf{q}[m] \mid t = l) \times P(t = l)$$

ID	Headache	Fever	Vomiting	Meningitis
1	true	true	false	false
2	false	true	false	false
3	true	false	true	false
4	true	false	true	false
5	false	true	false	true
6	true	false	true	false
7	true	false	true	false
8	true	false	true	true
9	false	true	false	false
10	true	false	true	true

Headache	Fever	Vomiting	Meningitis
true	true	false	?

ID	Headache	Fever	Vomiting	Meningitis
1	true	true	false	false
2	false	true	false	false
3	true	false	true	false
4	true	false	true	false
5	false	true	false	true
6	true	false	true	false
7	true	false	true	false
8	true	false	true	true
9	false	true	false	false
10	true	false	true	true

$$P(m \mid h, f, \neg v) = ?$$

$$P(\neg m \mid h, f, \neg v) = ?$$

$$P(m \mid h, f, \neg v) = \frac{\left(\begin{array}{l} P(h|m) \times P(f \mid h, m) \\ \quad \times P(\neg v \mid f, h, m) \times P(m) \end{array}\right)}{P(h, f, \neg v)}$$
$$= \frac{0.6666 \times 0 \times 0 \times 0.3}{0.1} = 0$$

$$P(\neg m \mid h, f, \neg v) = \frac{\left(\begin{array}{l} P(h|\neg m) \times P(f \mid h, \neg m) \\ \quad \times P(\neg v \mid f, h, \neg m) \times P(\neg m) \end{array}\right)}{P(h, f, \neg v)}$$
$$= \frac{0.7143 \times 0.2 \times 1.0 \times 0.7}{0.1} = 1.0$$

$$P(m \mid h, f, \neg v) = 0$$

$$P(\neg m \mid h, f, \neg v) = 1.0$$

- Υπάρχει κάτι παράξενο σε αυτά τα αποτελέσματα!

Κατάρα της Διαστατικότητας

Καθώς αυξάνεται ο αριθμός των περιγραφικών χαρακτηριστικών, αυξάνεται και ο αριθμός των πιθανών γεγονότων δέσμευσης. Κατά συνέπεια, απαιτείται εκθετική αύξηση του μεγέθους του συνόλου δεδομένων κάθε φορά που προστίθεται ένα νέο περιγραφικό χαρακτηριστικό, ώστε για οποιαδήποτε δεσμευμένη πιθανότητα να υπάρχουν αρκετά παραδείγματα στο σύνολο εκπαίδευσης που να ταιριάζουν στις συνθήκες, ώστε η προκύπτουσα πιθανότητα να είναι λογική.

- Η πιθανότητα ένας ασθενής που έχει πονοκέφαλο και πυρετό να έχει μηνιγγίτιδα θα πρέπει να είναι μεγαλύτερη από το μηδέν!
- Το σύνολο δεδομένων μας δεν είναι αρκετά μεγάλο $\rightarrow$ το μοντέλο μας κάνει υπερπροσαρμογή στα δεδομένα εκπαίδευσης.
- Οι έννοιες της δεσμευμένης ανεξαρτησίας και της παραγοντοποίησης μπορούν να μας βοηθήσουν να ξεπεράσουμε αυτό το ελάττωμα της τρέχουσας προσέγγισής μας.

- Αν η γνώση ενός γεγονότος δεν έχει καμία επίδραση στην πιθανότητα ενός άλλου γεγονότος, και *αντίστροφα*, τότε τα δύο γεγονότα είναι **ανεξάρτητα** μεταξύ τους.
- Αν δύο γεγονότα X και Y είναι ανεξάρτητα τότε:

$$P(X|Y) = P(X)$$
$$P(X, Y) = P(X) \times P(Y)$$

- Θυμηθείτε ότι όταν δύο γεγονότα είναι εξαρτημένα αυτοί οι κανόνες είναι:

$$P(X|Y) = \frac{P(X, Y)}{P(Y)}$$
$$P(X, Y) = P(X|Y) \times P(Y) = P(Y|X) \times P(X)$$

- Η πλήρης ανεξαρτησία μεταξύ γεγονότων είναι αρκετά σπάνια.
- Ένα πιο συνηθισμένο φαινόμενο είναι δύο ή περισσότερα γεγονότα να είναι ανεξάρτητα εφόσον γνωρίζουμε ότι έχει συμβεί ένα τρίτο γεγονός.
- Αυτό είναι γνωστό ως δεσμευμένη ανεξαρτησία.

- Για δύο γεγονότα, X και Y, που είναι δεσμευμένα ανεξάρτητα δεδομένης της γνώσης ενός τρίτου γεγονότος, εδώ του Z, ο ορισμός της πιθανότητας κοινού γεγονότος και της δεσμευμένης πιθανότητας είναι:

$$P(X|Y,Z) = P(X|Z)$$
$$P(X,Y|Z) = P(X|Z) \times P(Y|Z)$$

$$P(X|Y) = \frac{P(X,Y)}{P(Y)}$$
$$P(X,Y) = P(X|Y) \times P(Y)$$
$$= P(Y|X) \times P(X)$$

Χ και Υ είναι εξαρτημένα

$$P(X|Y) = P(X)$$
$$P(X,Y) = P(X) \times P(Y)$$

Χ και Υ είναι ανεξάρτητα

- Αν το γεγονός $t = l$ προκαλεί τα γεγονότα $\mathbf{q}[1], \ldots, \mathbf{q}[m]$, τότε τα γεγονότα $\mathbf{q}[1], \ldots, \mathbf{q}[m]$ είναι δεσμευμένα ανεξάρτητα μεταξύ τους δεδομένης της γνώσης του $t = l$, και ο ορισμός του κανόνα αλυσίδας μπορεί να απλοποιηθεί ως εξής:

$$
\begin{aligned}
&P(\mathbf{q}[1], \ldots, \mathbf{q}[m] \mid t = l) \\
&\qquad = P(\mathbf{q}[1] \mid t = l) \times P(\mathbf{q}[2] \mid t = l) \times \cdots \times P(\mathbf{q}[m] \mid t = l) \\
&\qquad = \prod_{i=1}^{m} P(\mathbf{q}[i] \mid t = l)
\end{aligned}
$$

- Χρησιμοποιώντας αυτό, μπορούμε να απλοποιήσουμε τους υπολογισμούς στο θεώρημα του Bayes, υπό την υπόθεση δεσμευμένης ανεξαρτησίας μεταξύ των περιγραφικών χαρακτηριστικών δεδομένου του επιπέδου l του χαρακτηριστικού-στόχου:

$$P(t = l \mid \mathbf{q}[1], \ldots, \mathbf{q}[m]) = \frac{\left(\prod_{i=1}^{m} P(\mathbf{q}[i] \mid t = l) \right) \times P(t = l)}{P(\mathbf{q}[1], \ldots, \mathbf{q}[m])}$$

Χωρίς δεσμευμένη ανεξαρτησία

$$P(X, Y, Z|W) = P(X|W) \times P(Y|X, W) \times P(Z|Y, X, W) \times P(W)$$

Με δεσμευμένη ανεξαρτησία

$$P(X, Y, Z|W) = \underbrace{P(X|W)}_{Factor1} \times \underbrace{P(Y|W)}_{Factor2} \times \underbrace{P(Z|W)}_{Factor3} \times \underbrace{P(W)}_{Factor4}$$

- Η κατανομή μικτής πιθανότητας για το σύνολο δεδομένων της μηνιγγίτιδας.

$$\mathbf{P}(H, F, V, M) = \begin{bmatrix} P(h, f, v, m), & P(\neg h, f, v, m) \\ P(h, f, v, \neg m), & P(\neg h, f, v, \neg m) \\ P(h, f, \neg v, m), & P(\neg h, f, \neg v, m) \\ P(h, f, \neg v, \neg m), & P(\neg h, f, \neg v, \neg m) \\ P(h, \neg f, v, m), & P(\neg h, \neg f, v, m) \\ P(h, \neg f, v, \neg m), & P(\neg h, \neg f, v, \neg m) \\ P(h, \neg f, \neg v, m), & P(\neg h, \neg f, \neg v, m) \\ P(h, \neg f, \neg v, \neg m), & P(\neg h, \neg f, \neg v, \neg m) \end{bmatrix}$$

- Αν υποθέσουμε ότι τα περιγραφικά χαρακτηριστικά είναι δεσμευμένα ανεξάρτητα μεταξύ τους δεδομένης της MENINGITIS, τότε χρειαζόμαστε να αποθηκεύσουμε μόνο τέσσερις παράγοντες:

$$Factor_1 : < P(M) >$$
$$Factor_2 : < P(h|m), P(h|\neg m) >$$
$$Factor_3 : < P(f|m), P(f|\neg m) >$$
$$Factor_4 : < P(v|m), P(v|\neg m) >$$

$$P(H, F, V, M) = P(M) \times P(H|M) \times P(F|M) \times P(V|M)$$

ID	Headache	Fever	Vomiting	Meningitis
1	true	true	false	false
2	false	true	false	false
3	true	false	true	false
4	true	false	true	false
5	false	true	false	true
6	true	false	true	false
7	true	false	true	false
8	true	false	true	true
9	false	true	false	false
10	true	false	true	true

- Υπολογίστε τους παράγοντες από τα δεδομένα.

$$Factor_1 : < P(M) >$$

$$Factor_2 : < P(h|m), P(h|\neg m) >$$

$$Factor_3 : < P(f|m), P(f|\neg m) >$$

$$Factor_4 : < P(v|m), P(v|\neg m) >$$

$$Factor_1 : < P(m) = 0.3 >$$
$$Factor_2 : < P(h|m) = 0.6666, P(h|\neg m) = 0.7413 >$$
$$Factor_3 : < P(f|m) = 0.3333, P(f|\neg m) = 0.4286 >$$
$$Factor_4 : < P(v|m) = 0.6666, P(v|\neg m) = 0.5714 >$$

$$Factor_1 : < P(m) = 0.3 >$$
$$Factor_2 : < P(h|m) = 0.6666, P(h|\neg m) = 0.7413 >$$
$$Factor_3 : < P(f|m) = 0.3333, P(f|\neg m) = 0.4286 >$$
$$Factor_4 : < P(v|m) = 0.6666, P(v|\neg m) = 0.5714 >$$

- Χρησιμοποιώντας τους παραπάνω παράγοντες, υπολογίστε την πιθανότητα MENINGITIS= *'true'* για το ακόλουθο ερώτημα.

HEADACHE	FEVER	VOMITING	MENINGITIS
true	true	false	?

$$P(m|h,f,\neg v) = \frac{P(h|m) \times P(f|m) \times P(\neg v|m) \times P(m)}{\sum_i P(h|M_i) \times P(f|M_i) \times P(\neg v|M_i) \times P(M_i)} =$$

$$\frac{0.6666 \times 0.3333 \times 0.3333 \times 0.3}{(0.6666 \times 0.3333 \times 0.3333 \times 0.3) + (0.7143 \times 0.4286 \times 0.4286 \times 0.7)} = 0.1948$$

$$Factor_1 : < P(m) = 0.3 >$$
$$Factor_2 : < P(h|m) = 0.6666, P(h|\neg m) = 0.7413 >$$
$$Factor_3 : < P(f|m) = 0.3333, P(f|\neg m) = 0.4286 >$$
$$Factor_4 : < P(v|m) = 0.6666, P(v|\neg m) = 0.5714 >$$

- Χρησιμοποιώντας τους παραπάνω παράγοντες, υπολογίστε την πιθανότητα MENINGITIS=*'false'* για το ίδιο ερώτημα.

HEADACHE	FEVER	VOMITING	MENINGITIS
true	true	false	?

$$P(\neg m|h,f,\neg v) = \frac{P(h|\neg m) \times P(f|\neg m) \times P(\neg v|\neg m) \times P(\neg m)}{\sum_i P(h|M_i) \times P(f|M_i) \times P(\neg v|M_i) \times P(M_i)} =$$

$$\frac{0.7143 \times 0.4286 \times 0.4286 \times 0.7}{(0.6666 \times 0.3333 \times 0.3333 \times 0.3) + (0.7143 \times 0.4286 \times 0.4286 \times 0.7)} = 0.8052$$

$$P(m|h, f, \neg v) = 0.1948$$

$$P(\neg m|h, f, \neg v) = 0.8052$$

- Όπως και πριν, η πρόβλεψη MAP θα ήταν MENINGITIS = *'false'*
- Οι εκ των υστέρων πιθανότητες δεν είναι τόσο ακραίες!

Τυπική Προσέγγιση: Ο Ταξινομητής Naive Bayes

Ταξινομητής Naive Bayes

$$\mathbb{M}(\mathbf{q}) = \underset{l \in levels(t)}{\operatorname{argmax}} \left(\prod_{i=1}^{m} P(\mathbf{q}[i] \mid t = l) \right) \times P(t = l)$$

Ο Naive Bayes εκπαιδεύεται απλά!

1. υπολογίστε τις εκ των προτέρων πιθανότητες για κάθε επίπεδο της μεταβλητής-στόχου
2. υπολογίστε τις δεσμευμένες πιθανότητες για κάθε χαρακτηριστικό δεδομένου κάθε επιπέδου της μεταβλητής-στόχου.

Πίνακας: Ένα σύνολο δεδομένων από πεδίο ανίχνευσης απάτης σε αιτήσεις δανείων.

ID	Credit History	Guarantor/ CoApplicant	Accomodation	Fraud
1	current	none	own	true
2	paid	none	own	false
3	paid	none	own	false
4	paid	guarantor	rent	true
5	arrears	none	own	false
6	arrears	none	own	true
7	current	none	own	false
8	arrears	none	own	false
9	current	none	rent	false
10	none	none	own	true
11	current	coapplicant	own	false
12	current	none	own	true
13	current	none	rent	true
14	paid	none	own	false
15	arrears	none	own	false
16	current	none	own	false
17	arrears	coapplicant	rent	false
18	arrears	none	free	false

$P(fr)$	$=$	0.3	$P(\neg fr)$	$=$	0.7
$P(\text{CH} = \text{'none'} \mid fr)$	$=$	0.1666	$P(\text{CH} = \text{'none'} \mid \neg fr)$	$=$	0
$P(\text{CH} = \text{'paid'} \mid fr)$	$=$	0.1666	$P(\text{CH} = \text{'paid'} \mid \neg fr)$	$=$	0.2857
$P(\text{CH} = \text{'current'} \mid fr)$	$=$	0.5	$P(\text{CH} = \text{'current'} \mid \neg fr)$	$=$	0.2857
$P(\text{CH} = \text{'arrears'} \mid fr)$	$=$	0.1666	$P(\text{CH} = \text{'arrears'} \mid \neg fr)$	$=$	0.4286
$P(\text{GC} = \text{'none'} \mid fr)$	$=$	0.8334	$P(\text{GC} = \text{'none'} \mid \neg fr)$	$=$	0.8571
$P(\text{GC} = \text{'guarantor'} \mid fr)$	$=$	0.1666	$P(\text{GC} = \text{'guarantor'} \mid \neg fr)$	$=$	0
$P(\text{GC} = \text{'coapplicant'} \mid fr)$	$=$	0	$P(\text{GC} = \text{'coapplicant'} \mid \neg fr)$	$=$	0.1429
$P(\text{ACC} = \text{'own'} \mid fr)$	$=$	0.6666	$P(\text{ACC} = \text{'own'} \mid \neg fr)$	$=$	0.7857
$P(\text{ACC} = \text{'rent'} \mid fr)$	$=$	0.3333	$P(\text{ACC} = \text{'rent'} \mid \neg fr)$	$=$	0.1429
$P(\text{ACC} = \text{'free'} \mid fr)$	$=$	0	$P(\text{ACC} = \text{'free'} \mid \neg fr)$	$=$	0.0714

Πίνακας: Οι πιθανότητες που χρειάζεται ένα μοντέλο πρόβλεψης Naive Bayes, υπολογισμένες από το σύνολο δεδομένων. Υπόμνημα συμβολισμού: FR=Fraudulent, CH=Credit History, GC = Guarantor/CoApplicant, ACC = Accomodation, T= *'true'*, F= *'false'*.

$P(fr)$	$=$	0.3	$P(\neg fr)$	$=$	0.7
$P(\text{CH} = \text{'none'} \mid fr)$	$=$	0.1666	$P(\text{CH} = \text{'none'} \mid \neg fr)$	$=$	0
$P(\text{CH} = \text{'paid'} \mid fr)$	$=$	0.1666	$P(\text{CH} = \text{'paid'} \mid \neg fr)$	$=$	0.2857
$P(\text{CH} = \text{'current'} \mid fr)$	$=$	0.5	$P(\text{CH} = \text{'current'} \mid \neg fr)$	$=$	0.2857
$P(\text{CH} = \text{'arrears'} \mid fr)$	$=$	0.1666	$P(\text{CH} = \text{'arrears'} \mid \neg fr)$	$=$	0.4286
$P(\text{GC} = \text{'none'} \mid fr)$	$=$	0.8334	$P(\text{GC} = \text{'none'} \mid \neg fr)$	$=$	0.8571
$P(\text{GC} = \text{'guarantor'} \mid fr)$	$=$	0.1666	$P(\text{GC} = \text{'guarantor'} \mid \neg fr)$	$=$	0
$P(\text{GC} = \text{'coapplicant'} \mid fr)$	$=$	0	$P(\text{GC} = \text{'coapplicant'} \mid \neg fr)$	$=$	0.1429
$P(\text{ACC} = \text{'own'} \mid fr)$	$=$	0.6666	$P(\text{ACC} = \text{'own'} \mid \neg fr)$	$=$	0.7857
$P(\text{ACC} = \text{'rent'} \mid fr)$	$=$	0.3333	$P(\text{ACC} = \text{'rent'} \mid \neg fr)$	$=$	0.1429
$P(\text{ACC} = \text{'free'} \mid fr)$	$=$	0	$P(\text{ACC} = \text{'free'} \mid \neg fr)$	$=$	0.0714

Credit History	Guarantor/CoApplicant	Accomodation	Fraudulent
paid	none	rent	?

$P(fr)$	$=$	0.3	$P(\neg fr)$	$=$	0.7
$P(\text{CH} = \text{'paid'} \mid fr)$	$=$	0.1666	$P(\text{CH} = \text{'paid'} \mid \neg fr)$	$=$	0.2857
$P(\text{GC} = \text{'none'} \mid fr)$	$=$	0.8334	$P(\text{GC} = \text{'none'} \mid \neg fr)$	$=$	0.8571
$P(\text{ACC} = \text{'rent'} \mid fr)$	$=$	0.3333	$P(\text{ACC} = \text{'rent'} \mid \neg fr)$	$=$	0.1429

$$\left(\prod_{k=1}^{m} P(\mathbf{q}[k] \mid fr) \right) \times P(fr) = 0.0139$$

$$\left(\prod_{k=1}^{m} P(\mathbf{q}[k] \mid \neg fr) \right) \times P(\neg fr) = 0.0245$$

Credit History	Guarantor/CoApplicant	Accomodation	Fraudulent
paid	none	rent	?

$P(fr)$	$=$	0.3	$P(\neg fr)$	$=$	0.7
$P(\text{CH} = \text{'paid'} \mid fr)$	$=$	0.1666	$P(\text{CH} = \text{'paid'} \mid \neg fr)$	$=$	0.2857
$P(\text{GC} = \text{'none'} \mid fr)$	$=$	0.8334	$P(\text{GC} = \text{'none'} \mid \neg fr)$	$=$	0.8571
$P(\text{ACC} = \text{'rent'} \mid fr)$	$=$	0.3333	$P(\text{ACC} = \text{'rent'} \mid \neg fr)$	$=$	0.1429

$$\left(\prod_{k=1}^{m} P(\mathbf{q}[k] \mid fr) \right) \times P(fr) = 0.0139$$

$$\left(\prod_{k=1}^{m} P(\mathbf{q}[k] \mid \neg fr) \right) \times P(\neg fr) = 0.0245$$

Credit History	Guarantor/CoApplicant	Accomodation	Fraudulent
paid	none	rent	*'false'*

Το μοντέλο γενικεύει πέρα από το σύνολο δεδομένων!

ID	Credit History	Guarantor/ CoApplicant	Accommodation	Fraud
1	current	none	own	true
2	paid	none	own	false
3	paid	none	own	false
4	paid	guarantor	rent	true
5	arrears	none	own	false
6	arrears	none	own	true
7	current	none	own	false
8	arrears	none	own	false
9	current	none	rent	false
10	none	none	own	true
11	current	coapplicant	own	false
12	current	none	own	true
13	current	none	rent	true
14	paid	none	own	false
15	arrears	none	own	false
16	current	none	own	false
17	arrears	coapplicant	rent	false
18	arrears	none	free	false
19	arrears	none	own	false
20	paid	none	own	false

Credit History	Guarantor/CoApplicant	Accommodation	Fraudulent
paid	none	rent	*'false'*

Σύνοψη

$$P(t|\mathbf{d}) = \frac{P(\mathbf{d}|t) \times P(t)}{P(\mathbf{d})} \tag{2}$$

- Ένας ταξινομητής Naive Bayes υποθέτει αφελώς ότι κάθε ένα από τα περιγραφικά χαρακτηριστικά ενός πεδίου είναι δεσμευμένα ανεξάρτητο από όλα τα υπόλοιπα περιγραφικά χαρακτηριστικά, δεδομένης της κατάστασης του χαρακτηριστικού-στόχου.
- Αυτή η υπόθεση, αν και συχνά λανθασμένη, επιτρέπει στο μοντέλο Naive Bayes να παραγοντοποιεί στο μέγιστο τη δομή αναπαράστασης που χρησιμοποιεί για το πεδίο.
- Παραδόξως, παρά την αφέλεια και τη δύναμη της υπόθεσης στην οποία βασίζεται, ένα μοντέλο Naive Bayes συχνά αποδίδει αρκετά καλά.