

Τεχνητή Νοημοσύνη I

Θεωρία λήψης αποφάσεων

Διδάσκων: Τσούρος Δήμος

Θεωρία λήψης αποφάσεων

Naive Bayes

- Ο *Naive Bayes* είναι ένας πιθανοτικός ταξινομητής που βασίζεται στο **Θεώρημα του Bayes** και υποθέτει **ανεξαρτησία** μεταξύ των χαρακτηριστικών.

Ιδέα: Για κάθε κατηγορία, υπολογίζουμε την πιθανότητα τα δεδομένα να ανήκουν σε αυτήν και επιλέγουμε τη μεγαλύτερη.

Υπολογίζουμε την πιθανότητα με βάση το θεώρημα του Bayes:

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

όπου C είναι η κατηγορία και $X = (x_1, x_2, \dots, x_n)$ τα χαρακτηριστικά.

Πλεονεκτήματα:

- Πολύ γρήγορος
- Απλός και αποδοτικός σε προβλήματα κειμένου

Μειονεκτήματα:

- Η υπόθεση ανεξαρτησίας συχνά δεν ισχύει πλήρως

Κανόνας απόφασης του Bayes

Έστω ότι είναι γνωστές

- οι εκ των προτέρων πιθανότητες $P(C_i)$ για κάθε κατηγορία C_i
- και οι υπό συνθήκη συναρτήσεις πυκνότητας πιθανότητας $P(X|C_i)$,
 - “πόσο πιθανό να έχω το δεδομένο X , θεωρώντας πως έχουμε την κατηγορία C_i »

Ο κανόνας απόφασης του Bayes είναι:

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

όπου C_i είναι η κατηγορία και $X = (x_1, x_2, \dots, x_n)$ τα χαρακτηριστικά, και

$$P(X) = \sum_{i=1}^n P(X|C_i)P(C_i)$$

Κανόνας απόφασης του Bayes

Ο τύπος απόφασης του Bayes μπορεί να περιγραφεί με λόγια ως εξής

$$a\ posteriori\ \text{πιθανότητα} = \frac{\text{πιθανοφανεια} * a\ prior\ \text{πιθανότητα κατηγορίας}}{\text{πιθανότητα γεγονότος}}$$

Ο κανόνας απόφασης του Bayes δηλώνει ότι:

- με την βοήθεια της παρατήρησης της τιμής του x
- είναι δυνατόν να μετατραπεί
 - η εκ των προτέρων πιθανότητα $P(C_i)$
 - στην εκ των υστέρων πιθανότητα $P(C_i|X)$,
- δηλαδή την πιθανότητα η κατάσταση της φύσης να είναι η C_i δεδομένου ότι έχει μετρηθεί η τιμή x για το δεδομένο.

Κανόνας απόφασης του Bayes

- $P(X|C_i)$ ονομάζεται *συνάρτηση πιθανοφάνειας* της κατηγορίας C_i σε σχέση με το X
- χρησιμοποιείται για να δηλώσει ότι:
 - εάν όλες οι υπόλοιποι παράμετροι είναι ίδιες,
 - η κατηγορία C_i για την οποία η $P(X|C_i)$ έχει μεγάλη τιμή έχει μεγαλύτερη πιθανότητα να είναι η σωστή κατηγορία.
- Να σημειωθεί ότι το **γινόμενο της πιθανοφάνειας και της εκ των προτέρων πιθανότητας** ($P(X|C_i) * P(C_i)$) είναι αυτό που καθορίζει την τιμή της εκ των υστέρων πιθανότητας.
- Ο παράγοντας $P(X)$ μπορεί να θεωρηθεί περισσότερο ως ένας παράγοντας *κανονικοποίησης*
 - που εγγυάται ότι το άθροισμα των εκ των υστέρων πιθανοτήτων θα ισούται με τη μονάδα.
 - Αλλά αφού είναι ο ίδιος διαιρέτης για όλες της κατηγορίες, **δεν επηρεάζει την απόφαση!**

Ο κανόνας απόφασης του Bayes:

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

Χρησιμοποιώντας τον κανόνα απόφασης του Bayes

Ο κανόνας απόφασης του Bayes θεωρεί γνωστες:

- Την εκ των προτέρων (a priori) πιθανότητα κάθε κατηγορίας
- Την συνάρτηση πιθανοφάνειας των τιμών των δεδομένων

Πως μπορούμε να τον χρησιμοποιήσουμε για πρόβλεψη όταν δεν γνωρίζουμε τις πραγματικές κατανομές;

Naïve Bayes: μάθε τις κατανομές από ένα σύνολο δεδομένων:

- $P(C_i)$: πιθανότητα εμφάνισης της κατηγορίας C_i στο σύνολο δεδομένων
- $P(X|C_i) = P(x_1|C_i) * \dots * P(x_k|C_i)$, με x_j να είναι η τιμή του X στο χαρακτηριστικό j
 - Θεωρεί πως τα χαρακτηριστικά x_j είναι ανεξάρτητα μεταξύ τους!
 - γι' αυτό Naïve!
- Θεωρώντας χαρακτηριστικά x_j που παίρνουν προκαθορισμένες κατηγορικές τιμές
 - $P(x_j|C_i)$: Το ποσοστό των φορών που εμφανίστηκε στα δεδομένα με κατηγορία C_i η τιμή του χαρακτηριστικού x_j
- Για κάθε νέο δεδομένο, πρόβλεψε την κατηγορία του με βάση αυτές τις κατανομές

Παράδειγμα

No	outlook	temperatur e	humidit y	windy	play
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	cool	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no

Naive Bayes

- $P(C_i)$: πιθανότητα εμφάνισης της κατηγορίας C_i στο σύνολο δεδομένων
 - $P(yes) = 9/14$
 - $P(no) = 5/14$
- Θεωρώντας χαρακτηριστικά x_j που παίρνουν προκαθορισμένες κατηγορικές τιμές
 - $P(x_j|C_i)$: Το ποσοστό των φορών που εμφανίστηκε στα δεδομένα με κατηγορία C_i η τιμή του χαρακτηριστικού x_j
 - Πρέπει να υπολογίσουμε την πιθανότητα **κάθε τιμής για κάθε κατηγορία**

Frequency Table

		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3



Likelihood Table

		Play Golf	
		Yes	No
Outlook	Sunny	3/9	2/5
	Overcast	4/9	0/5
	Rainy	2/9	3/5



		Play Golf	
		Yes	No
Humidity	High	3	4
	Normal	6	1

		Play Golf	
		Yes	No
Humidity	High	3/9	4/5
	Normal	6/9	1/5



		Play Golf	
		Yes	No
Temp.	Hot	2	2
	Mild	4	2
	Cool	3	1

		Play Golf	
		Yes	No
Temp.	Hot	2/9	2/5
	Mild	4/9	2/5
	Cool	3/9	1/5



		Play Golf	
		Yes	No
Windy	False	6	2
	True	3	3

		Play Golf	
		Yes	No
Windy	False	6/9	2/5
	True	3/9	3/5

Naive Bayes

Χρησιμοποιώντας τον κανόνα για πρόβλεψη σε νέα δεδομένα

Για κάθε νέο δεδομένο, πρόβλεψε την κατηγορία του με βάση αυτές τις κατανομές

Ο κανόνας απόφασης του Bayes:

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

Frequency Table

		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3



Likelihood Table

		Play Golf	
		Yes	No
Outlook	Sunny	3/9	2/5
	Overcast	4/9	0/5
	Rainy	2/9	3/5



		Play Golf	
		Yes	No
Humidity	High	3	4
	Normal	6	1

		Play Golf	
		Yes	No
Humidity	High	3/9	4/5
	Normal	6/9	1/5



		Play Golf	
		Yes	No
Temp.	Hot	2	2
	Mild	4	2
	Cool	3	1

		Play Golf	
		Yes	No
Temp.	Hot	2/9	2/5
	Mild	4/9	2/5
	Cool	3/9	1/5



		Play Golf	
		Yes	No
Windy	False	6	2
	True	3	3

		Play Golf	
		Yes	No
Windy	False	6/9	2/5
	True	3/9	3/5

Outlook	Temp	Humidity	Windy	Play
Rainy	Cool	High	True	?

$$P(Yes | X) = P(Rainy | Yes) \times P(Cool | Yes) \times P(High | Yes) \times P(True | Yes) \times P(Yes)$$

$$P(Yes | X) = 2/9 \times 3/9 \times 3/9 \times 3/9 \times 9/14 = 0.00529 \rightarrow 0.2 = \frac{0.00529}{0.02057 + 0.00529}$$

$$P(No | X) = P(Rainy | No) \times P(Cool | No) \times P(High | No) \times P(True | No) \times P(No)$$

$$P(No | X) = 3/5 \times 1/5 \times 4/5 \times 3/5 \times 5/14 = 0.02057 \rightarrow 0.8 = \frac{0.02057}{0.02057 + 0.00529}$$

Υπολογίστε τις πιθανότητες κάθε κατηγορίας για τα ακόλουθα

Outlook	Temp	Humidity	Windy
rainy	cool	high	FALSE
rainy	mild	normal	FALSE
sunny	hot	normal	FALSE

Naive Bayes – Παράδειγμα 2

Day	Discount	Free Delivery	Purchase
Weekday	Yes	Yes	Yes
Weekday	Yes	Yes	Yes
Weekday	No	No	No
Holiday	Yes	Yes	Yes
Weekend	Yes	Yes	Yes
Holiday	No	No	No
Weekend	Yes	No	Yes
Weekday	Yes	Yes	Yes
Weekend	Yes	Yes	Yes
Holiday	Yes	Yes	Yes
Holiday	No	Yes	Yes
Holiday	No	No	No
Weekend	Yes	Yes	Yes
Holiday	Yes	Yes	Yes
Holiday	Yes	Yes	Yes
Weekday	Yes	Yes	Yes
Holiday	No	Yes	Yes
Weekday	Yes	No	Yes
Weekend	No	No	Yes
Weekend	No	Yes	Yes
Weekday	Yes	Yes	Yes
Weekend	Yes	Yes	No
Holiday	No	Yes	Yes
Weekday	Yes	Yes	Yes
Holiday	No	No	No
Weekday	No	Yes	No
Weekday	Yes	Yes	Yes
Weekday	Yes	Yes	Yes
Holiday	Yes	Yes	Yes
Weekend	Yes	Yes	Yes

Frequency Table		Buy	
		Yes	No
Discount	Yes	19	1
	No	5	5

Frequency Table		Buy	
		Yes	No
Free Delivery	Yes	21	2
	No	3	4

Frequency Table		Buy	
		Yes	No
Day	Weekday	9	2
	Weekend	7	1
	Holiday	8	3

Υπολογίστε τις πιθανότητες κάθε κατηγορίας για τα ακόλουθα

Day	Discount	Free delivery
weekday	no	yes
weekday	yes	no
weekend	yes	yes

Naive Bayes – Παράδειγμα 2

Day	Discount	Free Delivery	Purchase
Weekday	Yes	Yes	Yes
Weekday	Yes	Yes	Yes
Weekday	No	No	No
Holiday	Yes	Yes	Yes
Weekend	Yes	Yes	Yes
Holiday	No	No	No
Weekend	Yes	No	Yes
Weekday	Yes	Yes	Yes
Weekend	Yes	Yes	Yes
Holiday	Yes	Yes	Yes
Holiday	No	Yes	Yes
Holiday	No	No	No
Weekend	Yes	Yes	Yes
Holiday	Yes	Yes	Yes
Holiday	Yes	Yes	Yes
Weekday	Yes	Yes	Yes
Holiday	No	Yes	Yes
Weekday	Yes	No	Yes
Weekend	No	No	Yes
Weekend	No	Yes	Yes
Weekday	Yes	Yes	Yes
Weekend	Yes	Yes	No
Holiday	No	Yes	Yes
Weekday	Yes	Yes	Yes
Holiday	No	No	No
Weekday	No	Yes	No
Weekday	Yes	Yes	Yes
Weekday	Yes	Yes	Yes
Holiday	Yes	Yes	Yes
Weekend	Yes	Yes	Yes

Frequency Table		Buy	
		Yes	No
Discount	Yes	19	1
	No	5	5

Frequency Table		Buy	
		Yes	No
Free Delivery	Yes	21	2
	No	3	4

Frequency Table		Buy	
		Yes	No
Day	Weekday	9	2
	Weekend	7	1
	Holiday	8	3

Πέρα από την πρόβλεψη των κατηγοριών νέων δεδομένων, μπορούμε να υπολογίσουμε την «σημασία» κάθε χαρακτηριστικού για την πρόβλεψη:

$$P(\text{No Buy}) = 6/30 = 0.2$$

$$P(\text{Weekday}) = 11/30 = 0.37$$

$$P(\text{Weekday} / \text{No Buy}) = 2/6 = 0.33$$

Βρείτε το $P(\text{No Buy} / \text{Weekday})$ χρησιμοποιώντας το θεώρημα Bayes,

$$P(\text{No Buy} / \text{Weekday}) = P(\text{Weekday} / \text{No Buy}) * P(\text{No Buy}) / P(\text{Weekday})$$

$$= (2/6 * 6/30) / (11/30) = 0.1818$$

Ελαχιστοποίηση κόστους με την θεωρία απόφασης του Bayes

Έστω $\{C_1, C_2, \dots, C_n\}$ το πεπερασμένο σύνολο η διαφορετικών κατηγοριών (καταστάσεων της φύσης) και $\{\alpha_1, \alpha_2, \dots, \alpha_m\}$ πεπερασμένο σύνολο των m πιθανών ενεργειών.

- Μπορεί οι πιθανές ενέργειες να είναι περισσότερες ή λιγότερες από τις κατηγορίες

Η συνάρτηση κόστους $\lambda(\alpha_j | C_i)$ περιγράφει το κόστος που αντιστοιχεί στην ενέργεια α_j όταν η κατάσταση της φύσης είναι η C_i .

- Έστω ότι παρατηρείται ένα συγκεκριμένο δεδομένο X και λαμβάνεται η ενέργεια α_j .
- Το **πραγματικό κόστος** της ενέργειας α_j θα είναι $\lambda(\alpha_j | C_i)$, με C_i την πραγματική κατηγορία του.
- Αφού η πιθανότητα η πραγματική κατηγορία να είναι η C_i είναι $P(C_i | X)$, το **αναμενόμενο κόστος** $R(\alpha_j | X)$ για την ενέργεια α_j θα είναι: $R(\alpha_j | X) = \sum_{i=1}^n \lambda(\alpha_j | C_i) * P(C_i | X)$
 - Το αναμενόμενο κόστος για μια ενέργεια ονομάζεται υπό συνθήκη **ρίσκο**
 - Για να ελαχιστοποιήσουμε το **ρίσκο** (*risk minimization*) επιλέγουμε την ενέργεια με το μικρότερο αναμενόμενο κόστος (υπό συνθήκη **ρίσκο**) $R(\alpha_j | X)$

Ελαχιστοποίηση κόστους με την θεωρία απόφασης του Bayes

Υπό συνθήκη ρίσκο: $R(\alpha_j|X) = \sum_{i=1}^n \lambda(\alpha_j|C_i) * P(C_i|X)$

- Για να ελαχιστοποιηθεί το συνολικό ρίσκο, υπολογίζουμε το υπό συνθήκη ρίσκο για $j = 1, \dots, m$ και στη συνέχεια επιλέγουμε την ενέργεια α_j για την οποία το $R(\alpha_j|X)$ είναι ελάχιστο.
- Το ελάχιστο συνολικό ρίσκο που προκύπτει καλείται ρίσκο του Bayes, συμβολίζεται με R^* , και είναι η βέλτιστη απόδοση που μπορεί να επιτευχθεί, με βάση τις πιθανοτικές κατανομές που γνωρίζουμε

Ελαχιστοποίηση κόστους στην κατηγοριοποίηση με 2 κατηγορίες

Στην περίπτωση αυτή η ενέργεια α_1 αντιστοιχεί στην απόφαση ότι η πραγματική κατάσταση της φύσης είναι η C_1 και η ενέργεια α_2 στην απόφαση ότι είναι η C_2 .

Συμβολίζουμε με $\lambda_{ji} = \lambda(\alpha_j|C_i)$ το κόστος όταν η ενέργειά μας είναι να προβλέψουμε C_j ενώ η πραγματική κατηγορία είναι η C_j .

Χρησιμοποιώντας την εξίσωση για το υπό συνθήκη ρίσκο

$(R(\alpha_j|X) = \sum_{i=1}^n \lambda(\alpha_j|C_i) * P(C_i|X))$ έχουμε:

- $R(\alpha_1|X) = \lambda_{11} * P(C_1|X) + \lambda_{12} * P(C_2|X)$
- $R(\alpha_2|X) = \lambda_{21} * P(C_1|X) + \lambda_{22} * P(C_2|X)$

Ο βασικός κανόνας απόφασης ελάχιστου ρίσκου είναι να αποφασίσουμε C_1 εάν:

$$\begin{aligned} R(\alpha_1|X) &< R(\alpha_2|X) \\ \lambda_{11} * P(C_1|X) + \lambda_{12} * P(C_2|X) &< \lambda_{21} * P(C_1|X) + \lambda_{22} * P(C_2|X) \\ (\lambda_{11} - \lambda_{21}) * P(C_1|X) &< (\lambda_{22} - \lambda_{12}) * P(C_2|X) \end{aligned}$$

Ελαχιστοποίηση κόστους στην κατηγοριοποίηση με 2 κατηγορίες

Ο βασικός κανόνας απόφασης ελάχιστου ρίσκου είναι να αποφασίσουμε C_1 εάν:

$$\begin{aligned} R(\alpha_1|X) &< R(\alpha_2|X) \\ \lambda_{11} * P(C_1|X) + \lambda_{12} * P(C_2|X) &< \lambda_{21} * P(C_1|X) + \lambda_{22} * P(C_2|X) \\ (\lambda_{11}-\lambda_{21}) * P(C_1|X) &< (\lambda_{22} - \lambda_{12}) * P(C_2|X) \end{aligned}$$

Μπορούμε να το φέρουμε σε εναλλακτική μορφή, η οποία προκύπτει από το λογικό συλλογισμό ότι $\lambda_{ii} \neq \lambda_{ij}$:

$$\begin{aligned} (\lambda_{11}-\lambda_{21}) * P(C_1|X) &< (\lambda_{22} - \lambda_{12}) * P(C_2|X) \\ (\lambda_{11}-\lambda_{21}) * P(X|C_1) * P(C_1) &< (\lambda_{22} - \lambda_{12}) * P(X|C_2) * P(C_2) \\ (\lambda_{21}-\lambda_{11}) * P(X|C_1) * P(C_1) &> (\lambda_{12} - \lambda_{22}) * P(X|C_2) * P(C_2) \\ \frac{P(X|C_1)}{P(X|C_2)} &> \frac{(\lambda_{12} - \lambda_{22})}{(\lambda_{21}-\lambda_{11})} * \frac{P(C_2)}{P(C_1)} \end{aligned}$$

Το δεξί μέρος της σύγκρισης είναι μια σταθερή τιμή για κάθε X , επομένως, ο κανόνας απόφασης του Bayes μπορεί να ερμηνευθεί ως απόφαση για C_1 εάν ο λόγος πιθανοφάνειας είναι μεγαλύτερος από μία **τιμή κατωφλίου**.

Ελαχιστοποίηση ρυθμού λάθους στην κατηγοριοποίηση με 2 κατηγορίες

Στα προβλήματα κατηγοριοποίησης, κάθε ενέργεια α_i ερμηνεύεται συνήθως ως η απόφαση ότι η πραγματική κατάσταση της φύσης είναι η C_i .

- Εάν απλώς επιθυμούμε να αποφεύγονται τα λάθη, πρέπει να χρησιμοποιήσουμε έναν κανόνα απόφαση ο οποίος να ελαχιστοποιεί την πιθανότητα λάθους, δηλαδή τον **ρυθμό λάθους**.

Η συνάρτηση κόστους για αυτήν την περίπτωση είναι η, όπως καλείται και στη βιβλιογραφία, *συμμετρική ή μηδέν-ένα συνάρτηση κόστους*:

$$\lambda(\alpha_i | C_j) = \begin{cases} 0, & \text{αν } i = j \\ 1, & \text{αν } i \neq j \end{cases}$$

Αυτή η συνάρτηση κόστους δεν αντιστοιχεί κανένα κόστος στις σωστές αποφάσεις ενώ αντιστοιχεί μοναδιαίο κόστος σε κάθε λανθασμένη απόφαση. Έτσι, όλα τα λάθη είναι ισοδύναμα από πλευράς κόστους.

Σε αυτή τη περίπτωση επιστρέφουμε στον αρχικό τύπο:

$$\frac{P(X|C_1)}{P(X|C_2)} > \frac{(1-0)}{(1-0)} * \frac{P(C_2)}{P(C_1)}$$
$$\frac{P(X|C_1)}{P(X|C_2)} > \frac{P(C_2)}{P(C_1)} \rightarrow P(X|C_1) * P(C_1) > P(X|C_2) * P(C_2) \rightarrow P(C_1|X) > P(C_2|X)$$

Ελαχιστοποίηση κόστους - Παράδειγμα

Παράδειγμα που θεωρούμε πως η απόφαση να αποφασίσουμε play είναι πιο κοστοβόρα αν είναι λάθος, γιατί μπορεί να τραυματιστούμε:

$$\lambda(play|C_j) = \begin{cases} 0, & \text{αν } C_j = play \\ 3, & \text{αν } C_j \neq play \end{cases}$$

$$\lambda(No\ play|C_j) = \begin{cases} 0, & \text{αν } C_j = No\ play \\ 1, & \text{αν } C_j \neq No\ play \end{cases}$$

Outlook	Temp	Humidity	Windy
rainy	cool	high	FALSE
rainy	mild	normal	FALSE
sunny	hot	normal	FALSE

Frequency Table

		Play Golf	
		Yes	No
Outlook	Sunny	3	2
	Overcast	4	0
	Rainy	2	3



Likelihood Table

		Play Golf	
		Yes	No
Outlook	Sunny	3/9	2/5
	Overcast	4/9	0/5
	Rainy	2/9	3/5

		Play Golf	
		Yes	No
Humidity	High	3	4
	Normal	6	1



		Play Golf	
		Yes	No
Humidity	High	3/9	4/5
	Normal	6/9	1/5

		Play Golf	
		Yes	No
Temp.	Hot	2	2
	Mild	4	2
	Cool	3	1



		Play Golf	
		Yes	No
Temp.	Hot	2/9	2/5
	Mild	4/9	2/5
	Cool	3/9	1/5

		Play Golf	
		Yes	No
Windy	False	6	2
	True	3	3



		Play Golf	
		Yes	No
Windy	False	6/9	2/5
	True	3/9	3/5