Naiv Bayes klassifier

Aalborg Intelligence

Bayes klassifikation

Her gennemgåes teorien for Bayes naive klassifikation, men først ser vi på ideen bag med et eksempel.

Vi ser på en person og ønsker at give bud på om vedkommende stemmer på rød eller blå blok. Vi har på forhånd oplysninger om en del personer og ønsker at bruge den viden derfra til at give det bedste bud på om personen stemmer på rød eller blå blok.

Her har vi følgende data der viser hvem der stemmer på rød og blå blok.

	alle	mænd	kvinder	ung	ældre	Sjælland	Jylland	andet
rød	51,85%	48,00%	55,00%	65,00%	47,47%	54,00%	49,90%	52,78%
blå	48,15%	52,00%	45,00%	35,00%	52,53%	46,00%	50,10%	47,22%
antal	10000	4500	5500	2500	7500	3000	4500	2500

Giv det bedste bud på hvilken blok en person stemmer på:

a) Hvis det er en tilfældig person.

b) Hvis det er en mand.

Fra skemaet med oplysninger kan det værre sværre at give et bud på hvad en ældre kvinde fra Sjælland vil stemme på da oplysningen om køn tyder på personen vil stemme på rød, mens information om det er en ældre tyder på personen vil stemme på blå.

Her kunne vi selvfølgelig løse problemet ved at få information for hver kombination af køn, aldersgruppe og bopæl.

Hvis vi her ser på kombinationer af køn, aldersgruppe og bopæl ses at dette giver $2 \cdot 2 \cdot 3 = 12$ kombinationer, og hvis vi i stedet havde set på om man svarer ja eller nej til 50 spørgsmål, vil man kunne få 2^{50} forskellige kombinationer af svar.

Hvis man ser på en person der har svaret på de 50 spørgsmål kan man her forvente at man i ens data kun har ganske få eller ingen personer der har svaret på fuldstændig samme måde, og der vil ikke være meget at bassere ens bud på.

Derfor ønsker vi en metode hvor ens bud udelukkende basseres på information svarende til det fra skemaet hvor vi ikke ser på kombinationerne.

Bayes klassifier

I det følgende indfører vi det nødvendige matematik og notation til Naive Bayes klassifikation. Først og fremmest inføre vi en stokastisk variabel Y som kan antage de værdier som svarer til vores forskellige forudsigelser/bud. Ved eksemplet vil

$$Y \in \{bl, rd\}.$$

Y skal være en diskret Stokastisk variabel med et bestemt antal mulige udfald, og der behøver nødvendigvis ikke kun at være to udfald.

Derudover indører vi en stokastisk Variabel $\mathbf{X}$ hvor de mulige udfald er alle kombinationer af informationer, her kan vi egentlig tænke X som en stokastisk vektor $\mathbf{X} = (X_1, X_2, ..., X_q)$, hvor man ved eksemplet kunne sige X_1:køn, X_2:aldersgruppe og X_3:bopæl, og et udfald kunne være $x = (kvinde, ldre, Sjlland)$.

For hvert udfald af $\mathbf{Y}$ ønsker vi at bestemme sandsynligheden for værdien y antages når vi allerede har observeret at $\mathbf{X} = x$.

Sandsynligheden vil vi skrive som

$$P(Y = y \mid \mathbf{X} = \mathbf{x}).$$

Denne notation og betydningen deraf ser vi snart på.

Vi kalder $P(Y = y \mid \mathbf{X} = \mathbf{x})$ en **posterior sandsynlighed** fordi den udtrykker sandsynligheden for Y **efter** (post) vi har informationen $\mathbf{x}$.

Det mest sandsynlige udfald for Y når vi har informationen $\mathbf{x}$ betegnes $C(\mathbf{x})$ og kaldes **Bayes klassifikation**.

Betinget sandsynlighed og uafhængighed

Først vender vi dog lige tilbage til notationen $P(Y = y \mid \mathbf{X} = \mathbf{x})$ som vi kaldte posterior sandsynlighed. Dette er egentlig det vi definerer som en **betinget sandsynlighed**, hvilket er grunden til notationen $P(Y = y \mid \mathbf{X} = \mathbf{x})$.

Givet to hændelser A og B så benyttes notationen $P(A \mid B)$ for sandsynligheden for at A sker når det er givet at B sker. Det læses derfor også som sandsynligheden for A givet B.

Så $P(Y = y \mid \mathbf{X} = \mathbf{x})$ er derved sandsynligheden for $Y = y$ når det er givet at $\mathbf{X} = \mathbf{x}$.

Et banalt eksempel kunne være at Y angiver antal ben på et givent dyr, mens $\mathbf{X}$ angiver dyrearten. Her er det oplagt at sandsynligheden for fire eller to ben afhænger af hvilken dyreart der er tale om.

Formelt defineres betinget sandsynlighed for to *hændelser* A og B:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

Udtrykket $P(A \cap B)$ i tælleren har vi sandsynligheden for *fælleshændelsen* mellem A og B, og i nævneren sørger vi for at man kun ser på de udfald hvor B er givet. (Vi siger også at nævneren *normaliserer* sandsynligheden i forhold til sandsynligheden for hændelsen B).

Eksempel med betinget sandsynlighed

Antag vi fokuserer på en almindelig terning med seks sider. Lad B være hændelsen at antal øjne er mindre eller lig 3, $B = \{\ ,\ ,\ \}$, og lad hændelsen A være udfald med ulige antal øjne, $A = \{\ ,\ ,\ \}$. Da kan vi nemt indse at $P(A) = 3/6 = 1/2$ samt ligeledes $P(B) = 1/2$ pga det symmetriske udfaldsrum. Ser vi imidlertid på den betingede sandsynlighed for at A indtræffer givet at B er indtruffet får vi $P(A \mid B)$. Først ser vi $A \cap B = \{\ ,\ ,\ \} \cap \{\ ,\ ,\ \} = \{\ ,\ \}$, hvilket igen pga det symmetriske sandsynlighedsfelt betyder at $P(A \cap B) = 2/6 = 1/3$. Efter at vi normaliserer sandsynligheden ud fra betingelsen om at B **er** indtruffet får vi

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{1/3}{1/2} = \frac{2}{3}.$$

At betinge med hændelsen B svarer i dette simple eksempel til at *indskrænke* udfaldet for A fra alle ulige øjne til dem som er mindre eller lig 3. Der er således tre mulige udfald i vores B-verden hvoraf to er ulige.

Stokastisk uafhængighed

Når to hændelser A og B siges at være uafhængige betyder det at

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

Hvis vi ser på udtrykket for $P(A \mid B)$ og antager at A og B er uafhængige, ser vi at

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \overset{\text{Uafh.}}{=} \frac{P(A)P(B)}{P(B)} = P(A).$$

Bayes sætning

En meget vigtig matematisk egenskab ved betinget sandsynlighed er muligheden for at ombytte rollerne i formlen, således vi kan udtrykke $P(B \mid A)$ ud fra vores viden om $P(A \mid B)$. Sætningen kaldes Bayes sætning (eller formel) og kan let vises