

Naiv Bayes klassifiser

Aalborg Intelligence

Bayes klassifikation

For at introducere teorien om Bayes naive klassifikation, vil vi starte med at se på et eksempel for at få en idé om, hvad Bayes klassifikation går ud på.

Vi vil se på en person, og vi ønsker at give et bud på om vedkommende stemmer på rød eller blå blok. Vi har på forhånd oplysninger om en del andre personer og ønsker at bruge den viden 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

Opgave 1

Brug tabellen ovenfor og 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 være svært 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, at det er en ældre person, tyder på personen vil stemme på blå. Endelig vil oplysningen om, at kvinden bor på Sjælland igen få os til at tænke, at hun stemmer på rød blok.

Her kunne vi selvfølgelig løse problemet ved at få information for hver kombination af køn, aldersgruppe og bopæl. Men det viser sig ikke at være en helt gangbar fremgangsmåde. Forklaringen følger her: Hvis vi ser på kombinationer af køn, aldersgruppe og bopæl vil det i dette eksempel give $2 \cdot 2 \cdot 3 = 12$ 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 måske slet ingen personer, der har svaret på fuldstændig samme måde, og der vil ikke være meget at basere ens bud på.

Derfor ønsker vi en metode, hvor vores bud på hvad en ny person vil stemme på udelukkende baseres på information svarende til det fra skemaet ovenfor, hvor vi ikke ser på alle de forskellige kombinationer. Det er *det* Naive Bayes klassifikation kan.

Bayes klassifier

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

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

Lidt mere generelt siger man, at Y skal være en diskret stokastisk variabel med et bestemt antal mulige udfald, og der behøver altså ikke nødvendigvis kun at være to udfald.

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

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 en posterior sandsynlighed. I sandsynlighedssregningen kalder vi det også for 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 $P(A \mid B)$ som sandsynligheden for, at A sker, når det er givet, at B er sket. 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 som:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \quad (1)$$

Udtrykket $P(A \cap B)$ i tælleren er sandsynligheden for *fælleshændelsen* mellem A og B , og i nævneren sørger vi for, at man kun ser på de udfald, hvor B er givet¹.

¹ Man siger også, at nævneren *normaliserer* sandsynligheden i forhold til sandsynligheden for hændelsen B .

Eksempel med betinget sandsynlighed

Lad os fokusere på en almindelig terning med seks sider. Lad B være hændelsen at antal øjne er mindre eller lig med 3. Det vil sige, at hændelsen B består af udfaldene: $B = \{1, 2, 3\}$. Lad hændelsen A være udfald med ulige antal øjne: $A = \{1, 3, 5\}$.

Da kan vi nemt indse, at

$$P(A) = 3/6 = 1/2$$

samt ligeledes at

$$P(B) = 1/2$$

på grund af det symmetriske udfaldsrum.

Ser vi imidlertid på den betingede sandsynlighed for at A indtræffer givet, at B allerede er indtruffet, får vi $P(A | B)$. Det svarer til sandsynligheden for at slå et ulige antal øjne, hvis vi allerede ved at antallet af øjne er mindre end eller lig med 3.

Først ser vi, at $A \cap B = \{1, 3\} \cap \{1, 2, 3\} = \{1, 3\}$, hvilket igen på grund af 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

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

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 end eller lig med 3. Der er således tre mulige udfald i vores B -verden hvoraf to er ulige.

Stokastisk uafhængighed

Man siger, at to hændelser A og B er uafhængige af hinanden, hvis

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

Hvis vi ser på udtrykket for $P(A | B)$ i (1) og antager, at A og B er uafhængige, ser vi at

$$P(A | B) = \frac{P(A \cap B)}{P(B)} \stackrel{\text{uafh.}}{=} \frac{P(A) \cdot P(B)}{P(B)} = P(A)$$

Med andre ord betyder det, at sandsynligheden for A givet B er den samme som sandsynligheden for A . Det vil sige, at oplysningen om, at B allerede er indtruffet, ikke ændrer på sandsynligheden for A . Information om B tilfører altså ikke noget nyt i forhold til information om A , og det giver derfor mening af sige, at A og B er uafhængige af hinanden.

Bayes sætning

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

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

Da $A \cap B = B \cap A$ får vi at

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

Her kan $P(A | B)$ isoleres

$$P(A | B) = \frac{P(B | A) \cdot P(A)}{P(B)}.$$

Vi kan altså ved at kende $P(B | A)$, $P(B)$ og $P(A)$ udtrykke den betingede sandsynlighed $P(A | B)$. Vi vender lige om lidt tilbage til hvad vi kan bruge det til.

Som sidste bemærkning er det væsentligt at understrege at $P(A | B) \neq P(B | A)$ med mindre $P(A) = P(B)$ jf formelen ovenfor. F.eks. er sandsynligheden for et tilfældigt dyr er en elefant givet dyret har fire ben ikke den samme som sandsynligheden for at dyret har fire ben givet at dyret er en elefant!.

Binær Bayes klassifier

Antag nu at Y kun kan antage to tilstande som ved eksemplet med *rød* eller *blå*. I dette tilfælde oversætter man ofte de to udfald til hhv 0 og 1, eller i visse sammenhænge -1 og $+1$.

For en Bayes klassifikation betyder det at $C(\mathbf{x}) = 0$ hvis $P(Y = 0 | \mathbf{X} = \mathbf{x}) > P(Y = 1 | \mathbf{X} = \mathbf{x})$ og ellers $C(\mathbf{x}) = 1$.

Dette kan også udtrykkes på anden vis:

$$P(Y = 0 | \mathbf{X} = \mathbf{x}) > P(Y = 1 | \mathbf{X} = \mathbf{x}) \Leftrightarrow \frac{P(Y = 0 | \mathbf{X} = \mathbf{x})}{P(Y = 1 | \mathbf{X} = \mathbf{x})} > 1.$$

Her antager vi dog lige at

$$P(Y = 1 | \mathbf{X} = \mathbf{x}) \neq 0.$$

Her vil vi benytte os af **Bayes sætning** til at se på hvordan denne brøk kan beregnes.

Bayes sætning i anvendelse

Vi bruger først Bayes sætning til at udtrykke $P(A | X)$ og $P(B | X)$.

$$P(A | X) = \frac{P(X | A)P(A)}{P(X)} \quad \text{og} \quad P(B | X) = \frac{P(X | B)P(B)}{P(X)},$$

Når vi bestemmer forholdet mellem $P(A | X)$ og $P(B | X)$ vil vi kunne slippe af med nævneren som de har til fælles.

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

hvor vi egentlig udnytter regnereglen om brøkers brøker.

Vi kan nu benytte dette ved udtrykket for en binær Bayes klassifier fra tidligere.

$$\frac{P(Y = 0 | \mathbf{X} = \mathbf{x})}{P(Y = 1 | \mathbf{X} = \mathbf{x})} = \frac{P(\mathbf{X} = \mathbf{x} | Y = 0)P(Y = 0)}{P(\mathbf{X} = \mathbf{x} | Y = 1)P(Y = 1)}.$$

Naiv Bayes klassifier

Ud fra udtrykket for forholdet mellem de to posterior sandsynligheder $P(Y = 0 \mid \mathbf{X} = \mathbf{x})$ og $P(Y = 1 \mid \mathbf{X} = \mathbf{x})$ kan vi se at der indgår to typer af sandsynligheder: $P(\mathbf{X} = \mathbf{x} \mid Y = y)$ og $P(Y = y)$. Disse benævnes hhv likelihood og prior sandsynligheder idet $P(\mathbf{X} = \mathbf{x} \mid Y = y)$ udtrykker *likelihooden* (tro-ligheden) for at observere $\mathbf{X} = \mathbf{x}$ givet $Y = y$. Omvendt er prior sandsynligheden $P(Y = y)$ et udtryk for forhåndssandsynligheden for at $Y = y$.

Vi kan sammenfatte det således til det såkaldte Posterior forhold:

$$\underbrace{\frac{P(Y = 0 \mid \mathbf{X} = \mathbf{x})}{P(Y = 1 \mid \mathbf{X} = \mathbf{x})}}_{\text{Posterior forhold}} = \underbrace{\frac{P(\mathbf{X} = \mathbf{x} \mid Y = 0)}{P(\mathbf{X} = \mathbf{x} \mid Y = 1)}}_{\text{Likelihood forhold}} \underbrace{\frac{P(Y = 0)}{P(Y = 1)}}_{\text{Prior forhold}}.$$

Hvis vi vender tilbage til vores spørgsmål om at stemme på blå eller rød blok og $x = (\textit{kvinde}, \textit{ung}, \textit{Sjælland})$, så udtrykker $P(\mathbf{X} = \mathbf{x} \mid Y = 0)$ eksempelvis sandsynligheden for en person er en ung kvinde fra Sjælland givet at personen stemmer på *rød blok*. Når man skal bestemme den sandsynlighed skal vi huske at det er sandsynligheden for det samlede udsagn med køn, alder og bopæl.

For at kunne beregne ovenstående sandsynligheder bliver vi nødt til at antage en *model*. Én af de simpleste modeller er at antage at køn, alder og bopæl er uafhængige af hinanden givet $Y = y$. Denne forsimplende antagelse har medvirket til metodens navn: *Naiv Bayes* eller *Uafhængig Bayes klassifikation*.

Det betyder ifølge vores tidligere definition af uafhængighed at

$$\begin{aligned} P(\mathbf{X} = \mathbf{x} \mid Y = y) &= P(X_1 = x_1, X_2 = x_2, \dots, X_q = x_q \mid Y = y) \\ &= P(X_1 = x_1 \mid Y = y) P(X_2 = x_2 \mid Y = y) \cdots P(X_q = x_q \mid Y = y) \\ &= \prod_{i=1}^q P(X_i = x_i \mid Y = y), \end{aligned}$$

hvor \prod -symbolet i sidste linje betyder at vi tager produktet af alle leddene på formen $P(X_i = x_i \mid Y = y)$ fra $i = 1$ op til q -

altså præcist det som står i linjen over. Det minder således om sum-tegnet

$$\sum_{i=1}^n x_i = x_1 + x_2 + \cdots + x_n,$$

men blot for gange i stedet for plus.

Posterior forholdet, score og vægte

Samler vi nu udtrykkene som indgår i vores posterior forhold får vi nedenstående udtryk:

$$\frac{P(Y = 0 \mid \mathbf{X} = \mathbf{x})}{P(Y = 1 \mid \mathbf{X} = \mathbf{x})} = \frac{P(Y = 0)}{P(Y = 1)} \prod_{i=1}^q \frac{P(X_i = x_i \mid Y = 0)}{P(X_i = x_i \mid Y = 1)},$$

hvor hver faktor på højre siden bidrager ligeligt til om observationen \mathbf{x} skal klassificeres som $Y = 0$ eller $Y = 1$.

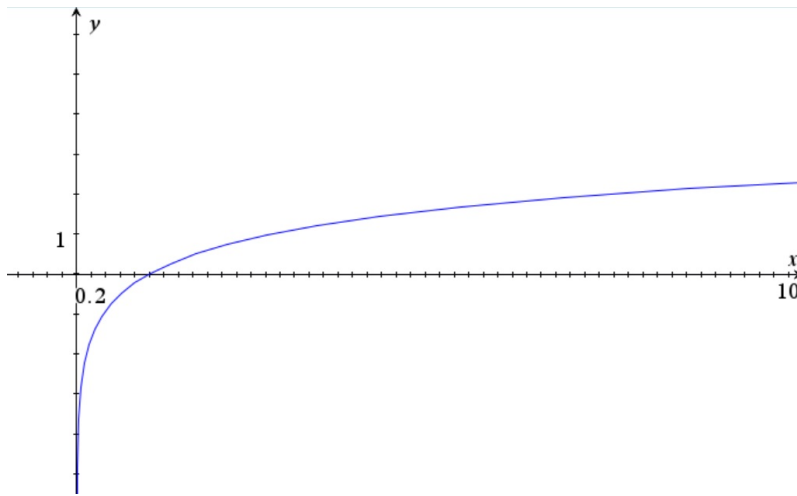
Når vi skal lave beregninger på computer baseret på data er det ofte væsentligt at tage højde for numerisk præcision. Alle tal på en computer skal *repræsenteres* af et endelig antal bits. Det betyder at visse tal (fx 1/3) bliver afrundet efter et vist antal decimaler. Derfor kan der opstå problemer når man enten ganger eller adderer meget små (eller store) tal sammen. For at undgå dette i udtrykket ovenfor, er det derfor tit en god idé at benytte sig af logaritmen på begge sider af lighedstegnet:

$$\ln \left(\frac{P(Y = 0 \mid \mathbf{X} = \mathbf{x})}{P(Y = 1 \mid \mathbf{X} = \mathbf{x})} \right) = \ln \left(\frac{P(Y = 0)}{P(Y = 1)} \right) + \sum_{i=1}^q \ln \left(\frac{P(X_i = x_i \mid Y = 0)}{P(X_i = x_i \mid Y = 1)} \right),$$

hvor vi brugte regneregler for logaritmen, $\ln(x \cdot y) = \ln(x) + \ln(y)$, gentagende gange.

Vi minder om at forholdet mellem $P(Y = 0 \mid \mathbf{X} = \mathbf{x})$ og $P(Y = 1 \mid \mathbf{X} = \mathbf{x})$ var af særlig interesse omkring værdien 1. Når deres forhold var 1 betyder det at de to klasser er lige sandsynlige givet \mathbf{x} . Endvidere, når forholdet var over 1 var $Y = 0$ mere sandsynlig end $Y = 1$, og når det var under 1 var det omvendte tilfældet.

Nu ses lige hvad effekten bliver af at benytte \ln . I figuren nedenfor er logaritmen plottet for $x \in [0, 10]$.



Vi ved at $\ln(1) = 0$, samt at for $x < 1$ er $\ln(x) < 0$ mens for $x > 1$ er $\ln(x) > 0$. Så når vi ser på

$$\ln \left(\frac{P(Y = 0 \mid \mathbf{X} = \mathbf{x})}{P(Y = 1 \mid \mathbf{X} = \mathbf{x})} \right)$$

bliver det vigtige om dette er positivt eller negativt.

Lad os indføre S som en *score* som er lig med logaritmen til posterior forholdet:

$$S = \ln \left(\frac{P(Y = 0 \mid \mathbf{X} = \mathbf{x})}{P(Y = 1 \mid \mathbf{X} = \mathbf{x})} \right)$$

Vi ved således at hvis $S > 0$ så klassificerer vi \mathbf{x} som $C(x) = 0$ og ellers $C(x) = 1$.

Indfører vi ligeledes bidragene til S som w_0 og $w_i(x_i)$ således

$$w_0 = \ln \left(\frac{P(Y = 0)}{P(Y = 1)} \right) \quad \text{og} \quad w_i(x_i) = \ln \left(\frac{P(X_i = x_i \mid Y = 0)}{P(X_i = x_i \mid Y = 1)} \right)$$

kan vi således skrive S som

$$S = w_0 + \sum_{i=1}^q w_i(x_i),$$

hvor det tydeliggøres at hvis $w_i(x_i) > 0$ understøtter bidraget fra den i 'te oplysning x_i at $Y = 0$ og ellers $Y = 1$. Denne egenskab gør at man også kan omtale $w_i(x_i)$ som en slags bevismæssig vægt.

Vægten w_0

Vi har altså set hvordan vi kan omskrive forholdet mellem posterior sandsynlighederne for de to klasser $Y = 0$ og $Y = 1$ til en sum af bidrag.

Det første led w_0 afhænger ikke af nogen information x_i og vi har tidligere omtalt disse sandsynligheder som *prior* sandsynligheder. Man kan sige at det er vores umiddelbare bud på hvad Y er uden at vi kender noget som helst til informationerne i \mathbf{x} .

Når vi går fra vores *model* som vi har udledt i det foregående til at skulle implementere den i en specifik anvendelse bliver vi derfor nødt til at estimere de parametre som indgår i modellen. For w_0 betyder det at vi skal estimere både $P(Y = 0)$ og $P(Y = 1)$. Her vil det være oplagt blot at estimere $P(Y = 0)$ og $P(Y = 1)$ ud fra ens træningsdata ved at bestemme andelen som stemmer på henholdsvis rød og blå, hvorefter vi kan bestemme

$$w_0 = \ln \left(\frac{P(Y = 0)}{P(Y = 1)} \right).$$

Her argumenteres lige lidt mere for dette valg. Vi ønsker at bestemme bedst muligt estimat for $p = P(Y = 0)$. Her kan vi se resultaterne fra datasættet som kommende fra et binomialforsøg med sandsynlighedsparameter p og antalsparameter n . Z er en stokastisk variabel der betegner antal der stemmer rød og fra binomialfordelingen ved vi at:

$$P(Z = r) = \binom{n}{r} p^r (1-p)^{n-r} = \frac{n!}{(n-r)!r!} p^r (1-p)^{n-r}.$$

Når vi skal estimere p ud fra data kan vi maksimere udtrykket ovenfor mht p ved at differentiere udtrykket og sætte lig 0. I stedet for at arbejde direkte med $P(Z = r)$ er det nemmere at arbejde med udtrykket for $\ln(P(Z = r))$ idet der gælder at $f(p) = P(Z = r)$ og $\ln(f(p))$ har maksimum ved samme p .

$$\frac{\partial}{\partial p} \ln(P(Z = r)) = \frac{\partial}{\partial p} \left(\ln \binom{n}{r} + r \ln p + (n-r) \ln(1-p) \right) = \frac{r}{p} - \frac{n-r}{1-p}.$$

Sætter vi ovenstående lig 0 og isolerer p får vi således $\hat{p} = r/n$, hvilket svarer til den andel af de n observationer som har $Y = 0$. Vi sætter en “*hat*” på p for at tydeliggøre at det er et estimat af p - og altså ikke den ukendte, *sande* værdi af p .

Vægtene w_i

De øvrige bidrag til S afhænger af den specifikke værdi af x_i . Det er altså her data for observationen vi ønsker at klassificere kommer ind i billedet. Her vil vægten $w_i(x_i)$, som bidrager til den samlede score S , afhænge af informationen x_i . Hvis $w_i(x_i)$ er mere eller mindre konstant for forskellige værdier af X_i betyder det at den i 'te information ikke er særlig informativ (og måske bør udelades fra modellen).

Ved hver information X_i estimeres $P(X_i = x_i \mid Y = y)$ ved at se på andelen af x_i blandt alle træningsdata med $Y = y$. Det er således på tilsvarende måde som for w_0 .

Når disse estimater er fundet kan man bestemme

$$w_i(x_i) = \ln \left(\frac{P(X_i = x_i \mid Y = 0)}{P(X_i = x_i \mid Y = 1)} \right).$$

Eksempel med rød/blå

Lad os se på eksemplet fra tidligere med at stemme på rød eller blå blok, hvor vi tænker på $Y = 0$ som stemme på rød og $Y = 1$ som stemme på blå.

Vi havde allerede følgende information fra træningsdata.

	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

Fra dette bestemmes først vægten w_0 ved

$$w_0 = \ln \left(\frac{P(Y = 0)}{P(Y = 1)} \right) = \ln \left(\frac{51,85\%}{48,15\%} \right) = 0,074.$$

For at kunne regne vores vægte w_i ud skal vi her have fat på hvor stor en del af stemmerne på rød der kommer fra henholdsvis mænd, kvinder, unge, ældre osv.

Vi tager beregningen for kvinder og starter med at finde $P(X_1 = \textit{kvinde} \mid Y = 0)$. Her ved vi at 55% af de 5500 kvinder stemte på rød blok, og samtidig ved vi at der var 5185 stemmer på rød blok. Derved får vi

$$P(X_1 = \textit{kvinde} \mid Y = 0) = \frac{55\% \cdot 5500}{5185} = 58,34\%.$$

Da X_1 kun kan antage værdierne kvinde og mand ved vi også at

$$P(X_1 = \textit{mand} \mid Y = 0) = 100\% - 58,34\% = 41,66\%.$$

På tilsvarende måde som finder vi finde $P(X_1 = \textit{kvinde} \mid Y = 1)$. Her ved vi at 45% af de 5500 kvinder stemte på blå blok, og samtidig ved vi at der var 4815 stemmer på blå blok. Derved får vi

$$P(X_1 = \textit{kvinde} \mid Y = 1) = \frac{45\% \cdot 5500}{4815} = 51,40\%$$

og

$$P(X_1 = \textit{mand} \mid Y = 1) = 100\% - 51,40\% = 48,60\%.$$

Alle tilsvarende sandsynligheder kan beregnes så man kan se hvor stor en andel af stemmerne på de to blokke der kommer fra hver gruppe.

	mænd	kvinder	ung	ældre	Sjælland	Jylland	andet
rød	41,66%	58,34%	31,34%	68,66%	31,24%	43,31%	25,45%
blå	48,60%	51,40%	18,17%	81,83%	28,66%	46,82%	24,52%

Nu kan vi bestemme vægtene $w_i(x_i)$. Her findes $w_1(\textit{kvinde})$ ved

$$w_1(\textit{kvinde}) = \ln \left(\frac{P(X_i = \textit{kvinde} \mid Y = 0)}{P(X_i = \textit{kvinde} \mid Y = 1)} \right) = \ln \left(\frac{58,34\%}{51,40\%} \right) = 0,1267.$$

Herunder ses vægtene for alle grupper:

	$prior(w_0)$	$w_1(mand)$	$w_1(kvinde)$	$w_2(ung)$	$w_2(ldre)$	$w_3(sjælland)$	$w_3(jylland)$	$w_3(øndet)$
vægt	0,074	-	0,127	0,545	-	0,086	-	0,037
		0,154			0,175		0,078	

Tidligere havde vi indset at når $w_i(x_i) > 0$ så understøtter bidraget fra den i 'te oplysning x_i at $Y = 0$ (altså stemme på rød). Derved ses at oplysningerne, kvinde, ung, sjælland og andet gør det mere sandsynligt med en stemme på rød, mens oplysningerne mand, ældre og Jylland gør det mere sandsynligt med en stemme på blå blok.

Vi kan nu beregnes Scoren S for en kvinde som er ældre og fra Sjælland, altså hvor $x = (kvinde, ldre, Sjælland)$.

$$\begin{aligned}
 S &= w_0 + \sum_{i=1}^q w_i(x_i) = w_0 + w_1(kvinde) + w_2(ldre) + w_3(Sjælland) \\
 &= 0,074 + 0,127 + (-0,176) + 0,086 = 0,11.
 \end{aligned}
 \tag{2}$$

Derved bliver forudsigelsen ud fra Bayes Naive metode at en ældre kvinde der bor på Sjælland med størst sandsynlighed stemmer på rød blok.