

# Sensitivitet og specificitet

Aalborg Intelligence

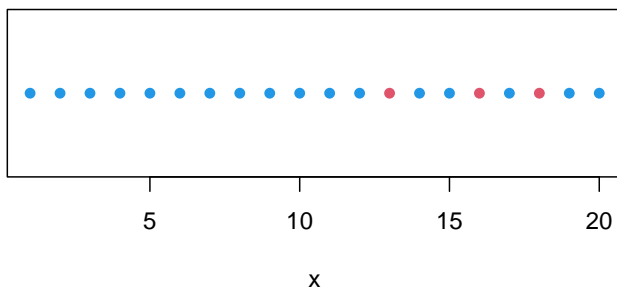
## Fejlklassifikationsraten

Når man skal vælge en god algoritme til at prædiktere klasser, har man brug for at kunne sammenligne, hvor godt forskellige algoritmer prædikterer. Man er derfor nødt til at have et mål for, hvor godt en algoritme prædikterer. Det mest oplagte er at tælle, hvor mange observationer algoritmen klassificerer forkert. Man kan så beregne, hvor stor en andel af alle observationerne, der klassificeres forkert. Denne andel kaldes *fejlklassifikationsraten*. Det er dog ikke altid det bedste mål at bruge. Det handler denne note om.

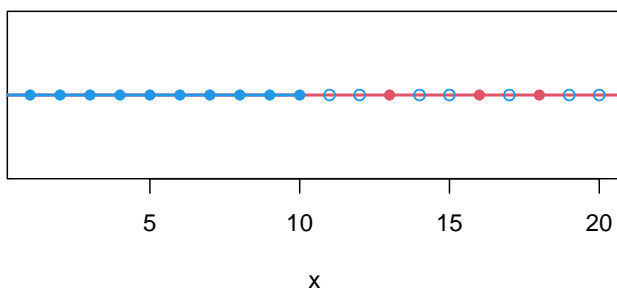
Som eksempel ser vi på et lille dataeksempel med 20 datapunkter, der kan have klasserne rød og blå, hvor rød er meget sjældnere end blå. Desuden er der målt en inputvariabel  $x$ . Vi ønsker at finde en algoritme, der kan prædiktere farven på en observation på baggrund af  $x$ . Datapunkternes klasser og  $x$ -værdier er angivet på figur 1. Ud fra figuren kunne det godt se ud til, at sandsynligheden for den røde klasse stiger, når  $x$  stiger. Der er dog også flest blå med meget høje  $x$ -værdier.

Mange prædiktionsalgoritmer benytter en tærskelværdi  $t$ , således at klassen prædikteres som rød, når  $x > t$ , og blå når  $x \leq t$ . På figur 2 ses et eksempel med  $t = 10$ . De første 10 observationer klassificeres korrekt til at være blå. De næste 10 observationer prædikteres røde, selv om kun 3 af dem faktisk er røde. Vi får altså 7 fejlklassifikationer i alt. Det giver en fejlklassifikationsrate på  $7/20 = 0.35$ .

Vi kan gøre det samme for forskellige værdier af  $t$  **ovenstående kunne laves til en slider** og tælle antallet af fejlklassifikationer. Resultatet ses i tabel 1.



Figur 1: Et lille dataeksempel med 20 datapunkter.



Figur 2: Klassifikation med  $t = 10$ . Observationer i det blå område prædikteres blå, mens observationer i det røde område prædikteres røde. Udfyldte cirkler angiver korrekte klassifikationer. Åbne cirkler angiver fejlklassifikationer.

Tabel 1: Fejlklassifikationer for forskellige værdier af  $t$

$t$	Antal fejl	Fejlklassifikationsrate
1	16	0.80
2	15	0.75
3	14	0.70
4	13	0.65
5	12	0.60
6	11	0.55
7	10	0.50
8	9	0.45
9	8	0.40
10	7	0.35
11	6	0.30
12	5	0.25
13	6	0.30
14	5	0.25
15	4	0.20
16	5	0.25
17	4	0.20
18	5	0.25
19	4	0.20
20	3	0.15

Den laveste fejlklassifikationsrate får vi ved at vælge  $t = 20$ , sådan at alle observationer prædikteres til at være blå. Men hvis vi er ude på at identificere de sjældne røde, så er sådan en test jo ikke meget værd. I ovenstående eksempel kunne man i stedet have valgt at sætte  $t = 12$ . Så får man ganske vist 5 fejlklassifikationer i stedet for 3. Til gengæld finder man alle de røde. Det virker som et mere fornuftigt valg i vores eksempel.

Hvordan vælger man så den bedste tærskelværdi? Det vil sige, hvordan finder man en god balance mellem ikke at lave for mange fejl og samtidig fange så mange som muligt fra den sjældne klasse? Her får vi brug for et mål for, hvor godt algoritmen prædikterer hver af de to klasser. Sensitivitet og specificitet er sådanne mål.

## Sensitivitet og specificitet

Lad os igen se på situationen, hvor vi har en prædiktionsalgoritme, der prædikterer klasserne rød og blå. For at få et overblik over, hvor godt algoritmen prædikterer, kan man lave en *confusion matrix* som skitseret på figur 3, hvor et positivt resultat svarer til klassen rød<sup>1</sup>. Diagonalen (de grønne celler) svarer til observationer, der er klassificeret korrekt. En god algoritme skal have så mange observationer som muligt på diagonalen.

		Sand klasse	
		Rød	Blå
Prædikteret klasse	Rød	Sand positiv	Falsk positiv
	Blå	Falsk negativ	Sand negativ

<sup>1</sup> Ordene positiv og negativ stammer fra medicin, hvor man bruger en test til at klassificere patienter som syge eller raske. En positiv test indikerer at patienten er syg. I andre sammenhænge svarer et positivt resultat til at man har prædikteret den sjældne klasse.

Figur 3: Illustration af confusion matrix

Hvis vi udelukkende er interesserede i, hvor god algoritmen er til at prædiktere den røde klasse, kan vi kigge på *sensitiviteten*. Dette er sandsynligheden for, at en observation, hvis sande farve er rød, faktisk bliver klassificeret som rød, altså

$$\text{Sensitivitet} = P(\text{En sand rød observation prædikteres som rød}) \quad (1)$$

Tilsvarende kan man måle, hvor god en algoritme er til at prædiktere den blå klasse ved at se på *specificiteten*. Dette er sandsynligheden for, at en sand blå observation faktisk bliver klassificeret som blå. Det kan udtrykkes som

$$\text{Specificitet} = P(\text{En sand blå observation prædikteres som blå})$$

Lad os se på eksemplet fra før, hvor vi sætter tærskelværdien til  $t = 15$ . Vi udfylder confusion matricen med antallet af observationer i hver celle.

Sensitiviteten beregnes som andelen af det samlede antal sande røde, der bliver prædikteret røde. Ved at se på første søjle i

		Sand klasse	
		Rød	Blå
Prædikeret klasse	Rød	2	3
	Blå	1	14

Figur 4: Confusion matrix med  $t = 15$

figur 4 finder vi, at der er  $2 + 1 = 3$  sande røde, hvoraf 2 bliver prædikeret røde. Det giver sensitiviteten

$$\text{Sensitivitet} = \frac{\text{Antal røde der prædikteres røde}}{\text{Antal røde}} = \frac{2}{3} = 0.667 \quad (2)$$

Tilsvarende kan vi beregne specificiteten ved at se på anden søjle.

$$\text{Specificitet} = \frac{\text{Antal blå der prædikteres blå}}{\text{Antal blå}} = \frac{14}{3 + 14} = 0.824$$

Algoritmen er altså bedst til at finde blå, da specificiteten er højere end sensitiviteten. Lad os prøve, om vi kan få højere sensitivitet med en anden værdi af  $t$ . Værdien  $t = 12$  var den mindste værdi, der kunne finde alle de røde. Det giver os confusion matrixen i figur 5.

		Sand klasse	
		Rød	Blå
Prædikeret klasse	Rød	3	5
	Blå	0	12

Figur 5: Confusion matrix med  $t = 12$

Vi beregner igen sensitiviteten og specificiteten

$$\text{Sensitivitet} = \frac{\text{Antal røde der prædikteres røde}}{\text{Antal røde}} = \frac{3}{3} = 1$$

og

$$\text{Specificitet} = \frac{\text{Antal blå der prædikteres blå}}{\text{Antal blå}} = \frac{12}{5 + 12} = 0.706$$

Vi ser altså, at prisen for at få en højere sensitivitet er en lavere specificitet.

Endelig kan vi prøve med værdien  $t = 20$ , som var den, der gav den laveste fejlklassifikationsrate. Denne værdi giver confusion matrixen i figur 6.

		Sand klasse	
		Rød	Blå
Prædikeret klasse	Rød	0	0
	Blå	3	17

Figur 6: Confusion matrix med  $t = 20$

Vi finder sensitiviteten

$$\text{Sensitivitet} = \frac{\text{Antal røde der prædikteres røde}}{\text{Antal røde}} = \frac{0}{3} = 0$$

og specificiteten

$$\text{Specificitet} = \frac{\text{Antal blå der prædikteres blå}}{\text{Antal blå}} = \frac{14}{17} = 1$$

Testen er altså rigtig god til at finde blå klasser, men elendig til at finde røde klasser.

### Sensitivitet og specificitet som betingede sandsynligheder

Sensitivitet og specificitet kan beskrives ved hjælp af betingede sandsynligheder. Lad  $A$  og  $B$  være to hændelser, således at  $B$  har positiv sandsynlighed  $P(B) > 0$ . Den betingede sandsynlighed for  $A$  givet  $B$  betegnes  $P(A|B)$  og er defineret som

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (3)$$

Her er  $A \cap B$  *fælleshændelsen*, det vil sige hændelsen at  $A$  og  $B$  forekommer samtidig. Vi fortolker  $P(A|B)$  som sandsynligheden for at hændelsen  $A$  indtræffer, hvis vi ved at hændelsen  $B$  er indtruffet. Dette giver mening i forhold til definitionen (3), idet brøken angiver, hvor stor en andel

af sandsynligheden for  $B$ , der udgøres af sandsynligheden for, at  $A$  indtræffer samtidig med  $B$ .

Lad os se på et eksempel, hvor vi slår to gange med en terning. Lad  $A$  være hændelsen, at vi slår to seksere, og lad  $B$  hændelsen, at den første terning viser seks. Da er  $P(A) = P(\text{to seksere}) = 1/36$  og  $P(B) = P(\text{første terning iser seks}) = 1/6$ . Intuitivt vil man forvente, at sandsynligheden for to seksere vokser, hvis den første terning viser en sekser. Det kan vi bekræfte ved hjælp af betingede sandsynligheder.

$$P(\text{to seksere} | \text{første terning viser seks}) = P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{1/36}{1/6} = \frac{1}{6}$$

Her udnyttede vi at  $A \cap B = A$ , da første terning er nødt til at vise seks for at vi kan få to seksere. Vi ser altså at

$$P(\text{to seksere} | \text{første terning viser seks}) = \frac{1}{6} \neq \frac{1}{36} = P(\text{to seksere})$$

Terningeksemplet viste et eksempel hvor  $P(A) \neq P(A|B)$ , altså hvor sandsynligheden for  $A$  ændrer sig, hvis vi ved, at  $B$  er indtruffet. Dette er ofte tilfældet. Nogle gange kan vi dog have at  $P(A|B) = P(A)$ , altså at vi ikke får nogen ny viden om sandsynligheden for  $A$  ud fra vores viden om  $B$ . I dette tilfælde siger vi, at  $A$  og  $B$  er *uafhængige*.

Sensitiviteten kan defineres formelt ved hjælp af betingede sandsynligheder som sandsynligheden for at få en rød prædiktion, *givet at den sande klasse er rød*, altså

$$\begin{aligned} \text{Sensitivitet} &= P(\text{Prædiktionen er rød} \mid \text{Den sande klasse er rød}) \\ &= \frac{P(\text{Den sande klasse er rød og prædiktionen er rød})}{P(\text{Den sande klasse er rød})} \end{aligned}$$

I praksis estimerer vi sandsynligheden for en sand rød som antallet af sande røde divideret med det samlede antal observationer. Sandsynligheden for, at en observation både er rød og klassificeres som rød, estimeres som antallet, der både er røde og klassificeres røde, divideret med det samlede antal observationer. Vi kan derfor estimere sen-

sitiviteten ved

$$\begin{aligned}\text{Sensitivitet} &= \frac{(\text{Antal røde der prædikteres røde})/(\text{Samlet antal})}{(\text{Antal røde})/(\text{Samlet antal})} \\ &= \frac{\text{Antal røde der prædikteres røde}}{\text{Antal røde}}\end{aligned}$$

Det var netop den formel, vi brugte til at udregne sensitiviteten i (1).

Specificiteten kan tilsvarende defineres som

$$\begin{aligned}\text{Specificitet} &= P(\text{Prædiktionen er blå} \mid \text{Den sande klasse er blå}) \\ &= \frac{P(\text{Den sande klasse er blå og prædiktionen er blå})}{P(\text{Den sande klasse er blå})}\end{aligned}$$

Man kan som ovenfor regne sig frem til, at specificiteten kan estimeres ved

$$\text{Specificitet} = \frac{\text{Antal blå der prædikteres blå}}{\text{Antal blå}}$$

## ROC-kurver

I praksis har man brug for at finde en god afvejning mellem sensitivitet og specificitet. Det kan afhænge af anvendelsen, hvor højt man vægter de to. Hvis man er ude på at diagnosticere en sjælden sygdom, er det umiddelbart vigtigst at sensitiviteten er høj, så man finder alle de syge. Dog er det problematisk, hvis specificiteten bliver for lav, da man så kommer til at diagnosticere mange raske som syge, hvilket kan medføre unødvendige undersøgelser og behandlinger for patienten.

Så hvordan vælger man en tærskelværdi, der giver en god afvejning mellem sensitivitet og specificitet? Som en hjælp kunne man udregne sensitivitet og specificitet for forskellige mulige værdier af  $t$ . Det er gjort i tabel 2.



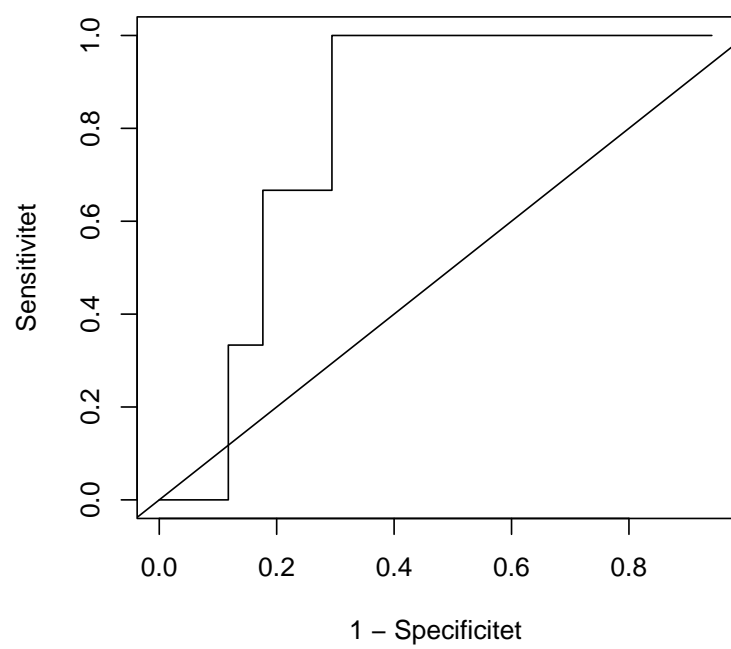
Tabel 2: Sensitivitet og specificitet for forskellige tærskelværdier

t	Sensitivitet	Specificitet
1	1	0.059
2	1	0.118
3	1	0.176
4	1	0.235
5	1	0.294
6	1	0.353
7	1	0.412
8	1	0.471
9	1	0.529
10	1	0.588
11	1	0.647
12	1	0.706
13	0.667	0.706
14	0.667	0.765
15	0.667	0.824
16	0.333	0.824
17	0.333	0.882
18	0	0.882
19	0	0.941
20	0	1

Man kan så gå ind i tabel 2 og lede efter et godt  $t$ , hvor både sensitivitet og specificitet er høj. En tabel som ovenfor bliver dog hurtigt uoverskuelig, hvis man har et stort datasæt. For at få overblik kan man i stedet vælge at tegne samhörrende værdier af sensitivitet og specificitet ind i et koordinatsystem. Traditionelt vælger man at have 1 – specificitet på  $x$ -aksen og sensitivitet på  $y$ -aksen. Den kurve der fremkommer, når punkterne forbindes, kaldes en *ROC-kurve*<sup>2</sup>. På figur 7 er ROC-kurven fra dataeksemplet i figur 1 indtegnet.

Vi vil gerne have både sensitivitet og specificitet til at være så tæt på 1 som muligt. Vi søger derfor et punkt på ROC-kurven, der ligger så tæt på punktet  $(0, 1)$  som muligt. Ud fra ROC-kurven kunne punktet  $(0.176, 0.667)$  ligne et godt bud. Ifølge tabellen svarer det til en tærskelværdi på  $t = 15$ .

<sup>2</sup> ROC står for *Receiver Operating Characteristic*



Figur 7: ROC-kurve for dataeksempel i figur [1](#)

På figur 7 er identitetslinjen  $y = x$  også indtegnet. Et punkt  $(p, p)$  på identitetslinjen svarer til en helt tilfældig prædiktionsalgoritme, hvor hver observation bliver klassificeret som rød med sandsynlighed  $p$  og blå med sandsynlighed  $1 - p$  uden at tage højde for værdien af  $x$ . For sådan en algoritme er sandsynligheden for at en sand rød prædikteres rød altså også  $p$ , så

$$\text{Sensitivitet} = P(\text{En sand rød prædikteres rød}) = P(\text{Rød prædiktion}) = p$$

Tilsvarende kan vi beregne specificiteten

$$\text{Specificitet} = P(\text{En sand blå prædikteres blå}) = P(\text{Blå prædiktion}) = 1 - p$$

og derfor

$$1 - \text{Specificitet} = 1 - (1 - p) = p$$

For sådan en test får vi altså

$$(1 - \text{specificitet}, \text{sensitivitet}) = (p, p)$$

Alt i alt viser dette, at punkterne på identitetslinjen svarer til helt tilfældig prædiktion. En prædiktionsalgoritme skal derfor helst give et punkt, der ligger over identitetslinjen. Ellers er den ikke bedre end et tilfældigt gæt.

## AUC

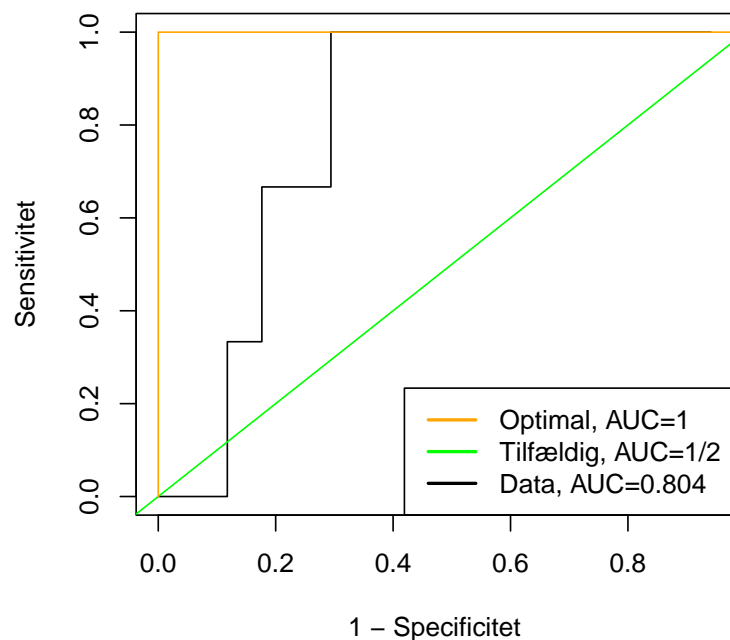
Hvis vi har brug for at sammenligne forskellige prædiktionsalgoritmer, kan det godt være svært at sammenligne deres fulde ROC-kurver. Det er nemmere at sammenligne et enkelt tal, der opsummerer, hvor god ROC-kuven er. Her kan AUC bruges.

Husk på, at ROC-kurven gerne skulle ligge så tæt op mod punktet  $(0, 1)$  og så langt over diagonalen som muligt. Vi kan derfor bruge *arealet under ROC-kurven*, også kaldet  $AUC$ <sup>3</sup>, som mål for hvor meget ROC-kurven er strakt opad mod  $(0, 1)$ .

<sup>3</sup> AUC står for *Area Under Curve*

Optimalt set skulle ROC-kurven stige lodret op til punktet  $(0, 1)$  og derefter fortsætte vandret over mod  $(1, 1)$  (den orange kurve på figur 8), svarende til at der er en  $t$ -værdi, der giver perfekt prædiktion. I denne situation er  $AUC = 1$ . Omvendt så vi, at identitetslinjen (den grønne kurve på figur 8) svarer til fuldstændig tilfældig prædiktion uden brug af  $x$ . Dette svarer til

$AUC = 1/2$ . En fornuftig algoritme skal således gerne have et  $AUC$  mellem  $1/2$  og  $1$ , hvor høje tal er bedst. I vores dataeksempel (den sorte kurve på figur 8) kan man udregne  $AUC = 0.804$ .



Figur 8: AUC for perfekt klassifikation (orange), tilfældig klassifikation (grøn) og vores dataeksempel (sort)

## Forskellige overvejelser

### Hvornår skal man bruge sensitivitet og specificitet?

Vi så i dataeksemplet fra figur 1, at fejlklassifikationsraten ikke egner sig som mål for, hvor godt en algoritme prædikerer, når en af klasserne er meget små. Her er det ofte en fordel at tillade flere fejlklassifikationer for at opnå en højere sensitivitet. Desuden kan det være en fordel at kigge på sensitivitet og specificitet i en situation, hvor man er mere interesseret i den ene klasse end i den anden. Det kunne fx være i forbindelse med test for sygdom under en epidemi, hvor det er vigtigere at finde alle

de syge, så de kan komme i karantæne, end at man undgår at sende raske i karantæne.

## Flere inputvariable

Ofte har man i praksis mere end én inputvariabel  $x$  at prædiktere ud fra. Lad os sige, at vi har målt variablene  $x_1, x_2, \dots, x_p$ . Mange algoritmer (blandt andet perceptronen, sigmoid neuroner, kunstige neurale netværk og logistisk regression **links, flere?**) laver prædiktionerne ud fra en vægtet sum af variablene

$$w_1x_1 + w_2x_2 + \dots + w_px_p$$

hvor  $w_1, w_2, \dots, w_p \in \mathbb{R}$  er konstanter. Man prædikterer så den ene klasse når

$$w_1x_1 + w_2x_2 + \dots + w_px_p > t$$

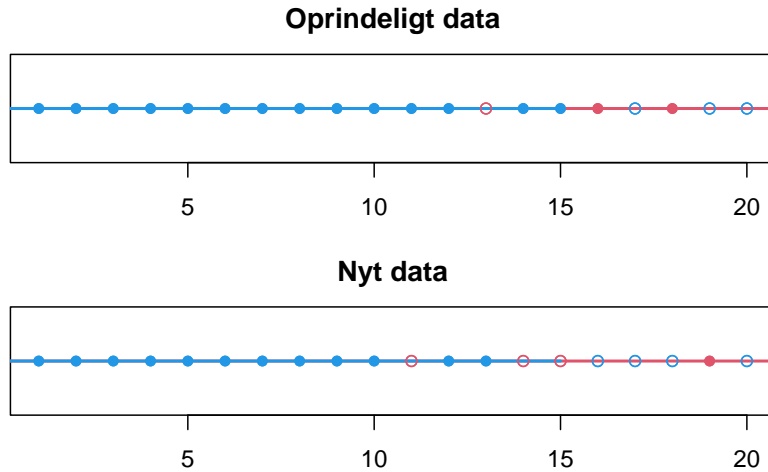
og den anden klasse når

$$w_1x_1 + w_2x_2 + \dots + w_px_p \leq t$$

hvor  $t$  er en passende tærskelværdi. Ofte bruger algoritmen som udgangspunkt  $t = 0$ . Som i tilfældet med én inputvariabel kan det dog give mening at vælge et andet  $t$  for at få bedre sensitivitet og specificitet. Igen kan man beregne confusion matrixen, sensitivitet og specificitet for forskellige værdier af  $t$  og tegne ROC-kurven for at finde et godt  $t$ . Vil man sammenligne flere algoritmer, kan man desuden beregne deres AUC ud fra ROC-kurven.

## Overfitting

I eksemplet fra figur 1 fandt vi, at  $t = 15$  virkede som et fornuftigt valg. Det var i hvert fald et  $t$ , der passede godt på det datasæt, vi havde. Det betyder dog ikke, at det er det  $t$ , der generaliserer bedst til nyt data. Lad os sige, at vi får et nyt datasæt og gerne vil bruge prædiktionsalgoritmen på det. På figur 9 ses et eksempel på, hvordan et nyt datasæt kunne se ud.



Figur 9: Det oprindelige data fra figur 1 og et nyt datasæt klassificeret ud fra  $t = 15$

Med  $t = 15$  får vi fejlklassificeret 3 ud af 4 røde i det nye datasæt, så sensitiviteten er  $1/4 = 0.25$ . Tilsvarende får vi fejlklassificeret 4 ud af 17 blå i det nye data, så specificiteten er  $13/17 = 0.765$ . Da vi brugte det oprindelige data fik vi sensitiviteten 0.667 og specificiteten 0.824. Både sensitivitet og specificitet er altså markant lavere for det nye data. Det sker, fordi  $t$  er valgt til at give høj sensitivitet og specificitet på lige præcis det oprindelige data. Det garanterer imidlertid ikke, at det passer lige så godt til nyt data. Vi siger at algoritmen er *overfittet* til det oprindelige data. Sensitivitet og specificitet giver altså ikke et retvisende mål for, hvor godt algoritmen prædikerer på nyt data. Du kan læse mere om overfitting [her link](#).

### Ekstra: Positiv og negativ prædiktiv værdi

Sensitivitet og specificitet bruges til at afgøre, hvor god en prædiktionsalgoritme er til at ramme rigtigt inden for hver klasse. Hvis vi for eksempel er ude på at prædiktere sygdom, så måler sensitiviteten sandsynligheden for, at en syg erklæres syg, og specificiteten måler sandsynligheden for, at en rask erklæres rask. En patient vil dog ofte være mere interesseret i det omvendte spørgsmål: “Jeg har fået en positiv test. Hvad er sandsynligheden for, at jeg faktisk er syg?” Det kan

lyde som næsten det samme, men det er faktisk et helt andet spørgsmål.

Lad os igen kigge på eksemplet med den røde og den blå klasse. Sensitiviteten var sandsynligheden for, at en sand rød observation bliver prædikeret som rød. Her tager vi altså udgangspunkt i, at *den sande klasse er rød* og kigger på sandsynligheden for, at observationen bliver klassificeret korrekt. I stedet kunne man kigge på sandsynligheden for, at en rød prædiktions faktisk betyder, at den sande klasse er rød. Her tager vi udgangspunkt i, at *prædiktationen er rød* og beregner sandsynligheden for, at den sande klasse er rød. Dette kaldes *den positive prædiktive værdi*, som altså er givet ved

Positiv prædiktiv værdi =  $P(\text{En rød prædiktation er faktisk rød})$

I praksis beregnes den positive prædiktive værdi ved formlen

$$\text{Positiv prædiktiv værdi} = \frac{\text{Antal røde prædiktioner som faktisk er røde}}{\text{Antal røde prædiktioner}}$$

Bemærk, at tælleren er den samme, som når vi beregner sensitivitet (2), mens nævneren er forskellig. Generelt vil sensitivitet og positiv prædiktiv værdi altså være forskellige tal.

Lad os igen se på vores lille dataeksempel med tærsklen  $t = 15$ , der gav anledning til confusion matrixen på figur 4. De røde prædiktioner findes i første række. Der er 5 røde prædiktioner i alt, hvoraf 2 faktisk er røde. Den positive prædiktive værdi kan udregnes til

$$\begin{aligned} \text{Positiv prædiktiv værdi} &= \frac{\text{Antal røde prædiktioner som faktisk er røde}}{\text{Antal røde prædiktioner}} \\ &= \frac{2}{5} = 0.4 \end{aligned} \quad (4)$$

Det betyder altså, at hvis vi har en rød prædiktation, så er sandsynligheden for, at den sande klasse er rød kun 0.4. Det er tilfældet på trods af, at både sensitivitet og specificitet var høje. Kort fortalt er grunden, at den røde klasse er så sjælden, at det er usandsynligt, at den sande klasse er rød, uanset om prædiktationen er rød eller blå. En mere præcis forklaring kan du finde i boksen nederst på siden.

Man kan selvfølgelig definere *den negative prædiktive værdi* tilsvarende.

Negativ prædiktiv værdi =  $P(\text{ En blå prædiktion er faktisk blå } )$

Den negative prædiktive værdi kan beregnes ved

$$\text{Negativ prædiktiv værdi} = \frac{\text{Antal blå prædiktionser som faktisk er blå}}{\text{Antal blå prædiktionser}}$$

Lad os igen se på vores lille dataeksempel med  $t = 15$  svarende til confusion matrixen på figur 4. For at finde den negative prædiktive værdi, bruger vi formlen

$$\begin{aligned}\text{Negativ prædiktiv værdi} &= \frac{\text{Antal blå prædiktionser som faktisk er blå}}{\text{Antal blå prædiktionser}} \\ &= \frac{14}{15} = 0.933\end{aligned}$$

Får man en blå prædiktion, kan man altså være 93.3% sikker på, at den er korrekt, mens man kun kunne være 40% sikker på en rød prædiktion.

Bemærk, at når vi beregner sensitivitet og specificitet, er det henholdsvis første og anden søjle i confusion matrixen, vi bruger, mens det er henholdsvis første og anden række i confusion matrixen, vi bruger til at beregne positiv og negativ prædiktiv værdi.

En væsentlig forskel på sensitivitet/specificitet og positiv/negativ prædiktiv værdi er, at sensitivitet og specificitet er faste egenskaber ved prædiktionsalgoritmen. De kan beregnes ved at teste algoritmen på en gruppe blå og en gruppe røde observationer og se, hvor ofte vi rammer plet. Positiv og negativ prædiktiv værdi afhænger derimod af hyppigheden af klasserne<sup>4</sup>. Det betyder for eksempel, at hvis man forsøger at prædiktere sygdom under en epidemi, så ændrer sandsynligheden for sygdom sig hele tiden, og det gør den positive og negative prædiktive værdi derfor også.

<sup>4</sup> De matematiske detaljer er givet i boksen nedenfor

### **Teori om sammenhængen mellem sensitivitet og positiv prædiktiv værdi**

Sensitiviteten var sandsynligheden for, at en sand rød blev



klassificeret som rød. Udtrykt ved betingede sandsynligheder var det

Sensitivitet =  $P(\text{Prædiktionen er rød} \mid \text{Den sande klasse er rød})$

Den positive prædiktive værdi kan tilsvarende udtrykkes ved hjælp af betingede sandsynligheder som

Positiv prædiktiv værdi =  $P(\text{Den sande klasse er rød} \mid \text{Prædiktionen er rød})$

De to formler minder meget om hinanden. Der er bare byttet om på de to hændelser i den betingede sandsynlighed.

Hvis  $A$  og  $B$  er to hændelser med  $P(A) > 0$  og  $P(B) > 0$ , så er  $P(A|B)$  og  $P(B|A)$  relateret via *Bayes' formel*

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

For at se, hvorfor det gælder, bruger vi først definitionen af  $P(A|B)$

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

Vi forlænger boksen med  $P(A)$  og bruger en brøkregerregel

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

Endelig bruger vi, at  $P(B|A) = \frac{P(A \cap B)}{P(A)}$ . Det giver

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

Vi har hermed bevist Bayes' formel.

Lader vi  $A = \{\text{Sand rød}\}$  og  $B = \{\text{Rød prædiktion}\}$  i Bayes' formel, får vi følgende sammenhæng mellem positiv prædiktiv værdi og sensitivitet

Positiv prædiktiv værdi =  $P(\text{Sand rød} \mid \text{Rød prædiktion})$

$$= P(\text{Rød prædiktion} \mid \text{Sand rød}) \cdot \frac{P(\text{Sand rød})}{P(\text{Rød prædiktion})}$$

$$= \text{Sensitivitet} \cdot \frac{P(\text{Sand rød})}{P(\text{Rød prädiktion})} \quad (5)$$

Formlen (5) viser, at hvis sandsynligheden for at tilhøre den røde klasse er meget lav i forhold til sandsynligheden for at lave en rød prädiktion, vil den positive prädiktive værdi være meget lavere end sensitiviteten.

Lad os se lidt nærmere på nævneren i (5), det vil sige  $P(B) = P(\text{Rød prädiktion})$ . Husk på at  $A = \{\text{Sand rød}\}$ . *Komplementærhændelsen* til  $A$  er hændelsen at  $A$  ikke indtræffer, og betegnes  $A^c$ . I vores tilfælde er  $A^c = \{\text{Ikke sand rød}\} = \{\text{Sand blå}\}$ . Hændelsen  $B$  at prädiktionen er rød kan opnås ved, at prädiktionen er rød, og den underliggende klasse er rød, svarende til  $B \cap A$ , eller ved at prädiktionen er rød, og den sande klasse er blå, svarende til  $B \cap A^c$ . Vi kan derfor beregne sandsynligheden for  $B$  som summen

$$P(B) = P(B \cap A) + P(B \cap A^c) \quad (6)$$

Vi bemærker nu, at definitionen af betinget sandsynlighed

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

kan omskrives til

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

På samme vis fås  $P(B|A^c)P(A^c) = P(B \cap A^c)$ . Dette kan vi indsætte i (6) og få

$$P(B) = P(B \cap A) + P(B \cap A^c) = P(B|A)P(A) + P(B|A^c)P(A^c)$$

Denne formel kaldes *loven om den totale sandsynlighed*.

Bruger vi loven om den totale sandsynlighed på formelen for positiv prädiktiv værdi (5), får vi

$$\begin{aligned} \text{Positiv prädiktiv værdi} &= \text{Sensitivitet} \cdot \frac{P(\text{Sand rød})}{P(\text{Rød prädiktion})} \\ &= \text{Sensitivitet} \cdot \frac{P(\text{Sand rød})}{P(\text{Rød prädiktion}|\text{Sand rød})P(\text{Sand rød}) + P(\text{Rød prädiktion}|\text{Sand blå})P(\text{Sand blå})} \end{aligned}$$

$$= \text{Sensitivitet} \cdot \frac{P(\text{Sand rød})}{\text{Sensitivitet} \cdot P(\text{Sand rød}) + (1 - \text{Specificitet}) \cdot P(\text{Sand blå})} \quad (7)$$

Vi ser, at hvis specificiteten ikke er meget høj, og sandsynligheden for sand rød er lav (og dermed sandsynligheden for sand blå høj), så er tælleren i (7) lille i forhold til nævneren. Den positive prædiktive værdi vil derfor være væsentligt lavere end specificiteten. Det var det, der skete i eksemplet (4).

Formlen (7) viser desuden, at den positive prædiktive værdi afhænger af ikke bare sensitivitet og specificitet, men også af fordelingen mellem de to klasser (altså sandsynligheden for sand rød og sand blå). Hvis fordelingen ændrer sig, så den røde klasse for eksempel bliver mere sandsynlig, så ændrer den positive prædiktive værdi sig også.