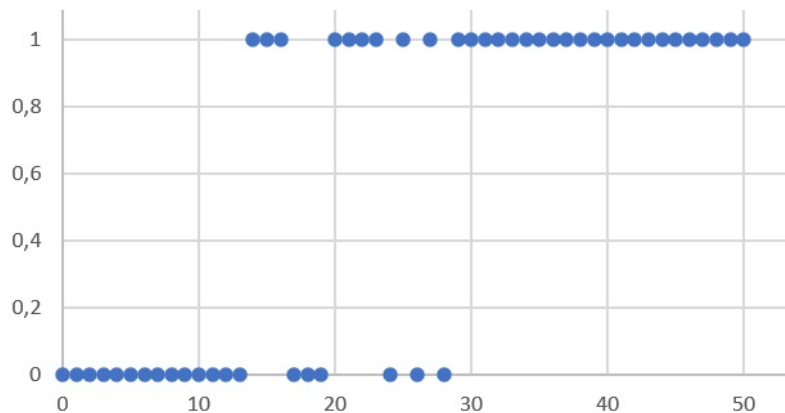


# Logistisk klassifikation

Aalborg Intelligence

## Ideen bag logistisk klassifikation

Lad os forestille os at vi ønsker at se på hvor hurtigt det går før en befolkning får immunitet mod en sygdom. Her tænker vi at det er sådan, at få fra starten af har sygdom/immunitet med at der på et tidspunkt sker en kæmpe udvikling og alle med tiden bliver immune. For at få styr over udvikling kunne man for hver time der går tage en test af en tilfældig person og se om personen er imun eller ikke er imun. Derved får man et datasæt med punkter  $(x,y)$  hvor  $x$  er tid(dage) og  $y$  kunne være 0 svarende til ikke imun eller 1 svarende til imun. Her kunne man forestille sig et datasæt som plottet herunder.

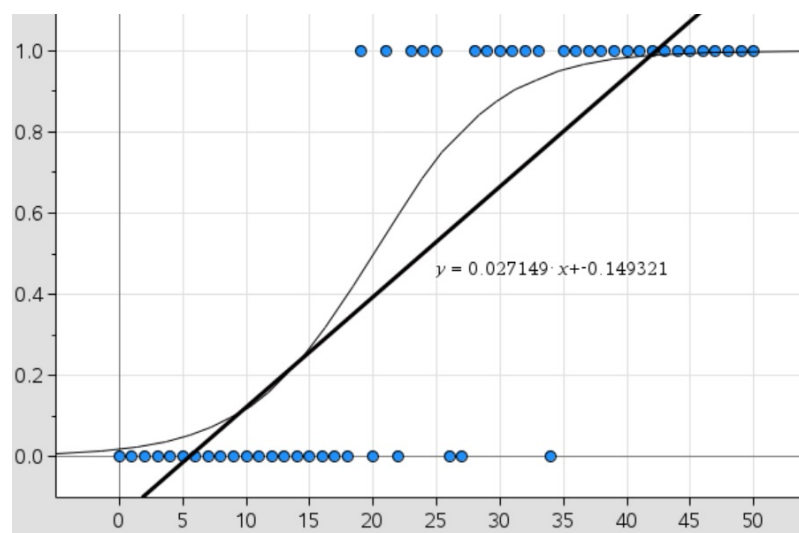


Figur 1: Datasæt hvor  $y=1$  svarer til imun

I stedet for at se på alle personer udtages nogen, og vi kunne være interesseret i på baggrund af datasættet at fastslå hvad sandsynligheden er for at man er imun på et givent tidspunkt.

Dette kunne ses som en funktion  $p(x)$  som vi derved er interesseret i at få indsigt i. Det er desværre ikke helt så let, og vi bliver nødt til at lave nogle antagelser inden vi kan komme med fornuftige forudsigelser.

Vi kunne starte med at se på om en lineær model ville fungere.



Figur 2: Datasæt med lineær regression og s-formet-funktion

Det er tydeligt at det ikke fungerer med en lineær funktion. (mere snak om hvorfor, her også fokus på problem med værdimængde som ikke er mellem 0 og 1).

Ved figuren er der også lagt en s-formet kurve ind med den egenskab at værdimængden er  $]0, 1[$  som umiddelbart ser mere fornuftig ud.

Istedet for at se på sandsynligheder direkte kan man se på to andre begreber, nemlig det som betegnes som "odds" som er forholdet mellem sandsynligheder for noget sker i forhold til sandsynligheden for det ikke sker, altså

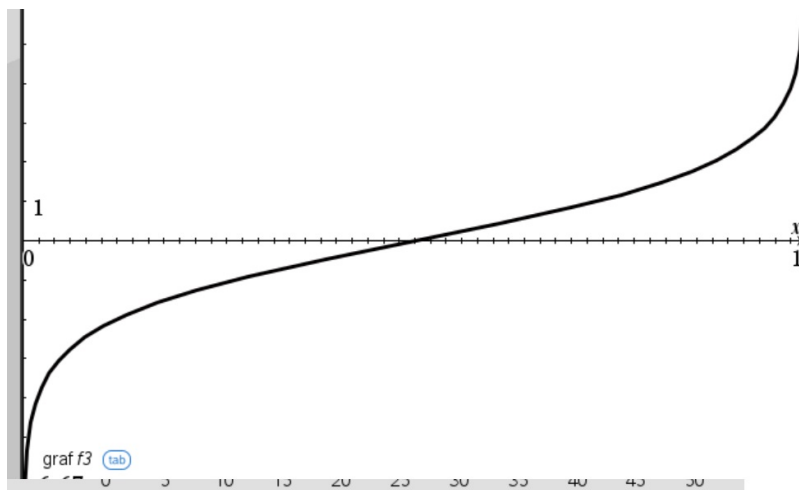
$$\frac{p(x)}{1 - p(x)}.$$

(note : mere teori om odds)

Desuden ses på den såkaldte logit-funktion som blot er den naturlige logaritme af odds, altså

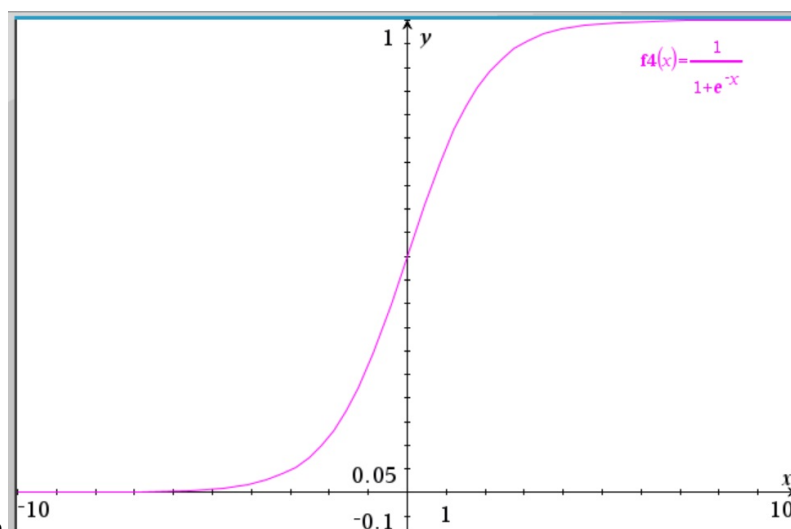
$$\ln \left( \frac{p(x)}{1 - p(x)} \right).$$

Hvis vi her ser dette som en funktion af  $p(x)$  med definitions-mængde mellem 0 og 1 ( $p(x)$  skal jo opfattes som sandsynlighed) ser vi at det giver følgende graf.



Det er en voksende funktion som derved har en invers, som man kan vise er givet ved

$$\frac{1}{1 + e^{-y}}.$$



Her ses grafen for denne

Denne graf ser mere fornuftig ud i forhold til vores data og den model vi ønsker. Vi kan se at det passer fint med at værdimængden  $]0, 1[$ , men selvfølgelig skulle der være mulighed for justering af hvor hurtig væksten kommer til at blive og en form for vandret forskyd.

Hvis man har en funktion  $f(x)$  kan væksten speedes op med en faktor  $k$  ved at se på  $f(k \cdot x)$ . Hvis man f.eks. ønsker at forskyde med  $h$  i  $x$ -retningen kan det gøres med funktionen  $f(k \cdot (x - h)) = f(k \cdot x - k \cdot h)$ . Her svarer  $k \cdot x - k \cdot h$  egentlig bare til en lineær funktion.

Derved ville det give mening at benytte modellen

$$\ln \left( \frac{p(x)}{1 - p(x)} \right) = a \cdot x + b$$

som netop er den model der anvendes ved logistisk klassifikation.

(evt. mere snak om funktionen  $\frac{1}{1+e^{-y}}$ . : Sigmond og bedre koblinger mellem funktionerne så ideen er mere klar)

## Funktionen $p(x)$

Vi ved nu at vi arbejder med følgende model

$$\ln \left( \frac{p(x)}{1 - p(x)} \right) = a \cdot x + b$$

Lad os se på hvordan man får et udtryk for  $p(x)$ . Vi starter med at skifte fortegn på begge sider af lighedstegnet.

$$-\ln \left( \frac{p(x)}{1 - p(x)} \right) = -a \cdot x - b.$$

Ved at benytte regnereglen  $\ln(a/b) = \ln(a) - \ln(b)$  frem og tilbage på venstresiden sammen med fortegnet fås

$$\ln \left( \frac{1 - p(x)}{p(x)} \right) = -a \cdot x - b.$$

Ved at benytte at  $e^x$  er invers til  $\ln(x)$  fås at

$$\frac{1-p(x)}{p(x)} = e^{-a \cdot x - b}$$

Her kan brøken deles op

$$\frac{1}{p(x)} - 1 = e^{-a \cdot x - b}$$

Her kan vi efter at have adderet med 1 få isoleret  $p(x)$  ved at gange med  $p(x)$  og efter følgende dividere med  $e^{-a \cdot x - b}$  hvorved vi ender ud med at

$$p(x) = \frac{1}{1 + e^{-a \cdot x - b}}.$$

Hvis vi forlænger med  $e^{a \cdot x + b}$  kan  $p(x)$  også skrives som

$$p(x) = \frac{e^{a \cdot x + b}}{1 + e^{a \cdot x + b}}.$$

## Maksimum Likelihood

Lad os starte med at forstille os at vi allerede kender funktionen  $p(x)$ . Derved kan vi beregne sandsynligheden for at vi får de givne observationer. Det  $i$ 'te punkt fra datasættet betegnes med  $(x_i, y_i)$  og antallet af observationer betegnes med  $n$ . For en observation  $(x_i, y_i)$  kender vi tidspunktet (svarende til  $x$ -værdien) men  $y$ -værdien afhænger af den tilfældigt udvalgte person. Sandsynligheden for at  $y$ -værdien antager værdien 1 er  $p(x)$ . Derved ved vi også at sandsynligheden for at  $y$ -værdien antager værdien 0 er  $1 - p(x)$ .

Vi antager at der er uafhængighed mellem vores observationerne hvorved sandsynligheder kan ganges sammen. Dette kan skrives ved hjælp af produkt-tegnet hvor vi ved det ene produkt indikerer at vi kun skal benytte de observationer hvor  $y_i = 1$  og ved den anden kun skal benytte dem hvor  $y_i = 0$ . Vi får følgende sandsynlighed, her betegnet med  $L$ , hørende til vores datasæt

$$L = \prod_{i: y_i=1} p(x_i) \cdot \prod_{i: y_i=0} (1 - p(x_i)).$$

(note med læsning af  $i : y_i = 1$ ).

Ved en model skal det selvfølgelig være sådan at den er tilpasset sådan at dette udtryk er størst muligt da modellen derved forudsiger at det observerede datasæt er mest sandsynligt.

Istedet for at se på netop dette udtryk tages  $\ln$  af udtrykket. Ved at benytte at  $\ln(a \cdot b) = \ln(a) + \ln(b)$  adskillige gange får vi at

$$\ln(L) = \sum_{i:y_i=1} \ln(p(x_i)) + \sum_{i:y_i=0} \ln(1 - p(x_i))$$

Da funktion  $f(x) = \ln(x)$  er en voksende funktion, vil en optimering af  $\ln(L)$  og  $L$  være det samme.

Nu ønsker vi egentlig at få udtrykt  $L$  hvor vi kun benytter en sum istedet for to. Til dette indser vi lige at

$$p(x_i)^{y_i} = \begin{matrix} 1 & \text{hvis} & y_i = 0 \\ p(x_i) & \text{hvis} & y_i = 1 \end{matrix}$$

og

$$(1 - p(x_i))^{1-y_i} = \begin{matrix} (1 - p(x_i)) & \text{hvis} & y_i = 0 \\ 1 & \text{hvis} & y_i = 1 \end{matrix}$$

Ved et produkt med flere faktorer har en faktor 1 selvfølgelig ikke nogen betydning, og vi får at  $L$  kan skrives som

$$L = \prod_i^n p(x_i)^{y_i} \cdot (1 - p(x_i))^{1-y_i}.$$

Ved at benytte at  $\ln(a \cdot b) = \ln(a) + \ln(b)$  og  $\ln(a^k) = k \cdot \ln(a)$  adskillige gange får vi at

$$\ln(L) = \sum_{i=1}^n y_i \cdot \ln(p(x_i)) + (1 - y_i) \cdot \ln(1 - p(x_i)).$$

Ved at ophæve parentesen fås

$$\ln(L) = \sum_{i=1}^n y_i \cdot \ln(p(x_i)) + \ln(1 - p(x_i)) - y_i \cdot \ln(1 - p(x_i)).$$

I to af ledene indenfor sumtegnet har vi  $y_i$  som en faktor, og  $y_i$  sættes udenfor parentes

$$\ln(L) = \sum_i^n \ln(1 - p(x_i)) + y_i \cdot (\ln(p(x_i)) - \ln(1 - p(x_i))).$$

Ved hjælp af logaritme-regnereglen  $\ln(a/b) = \ln(a) - \ln(b)$  får vi

$$\ln(L) = \sum_{i=1}^n \ln(1 - p(x_i)) + y_i \cdot \ln\left(\frac{p(x_i)}{1 - p(x_i)}\right).$$

Her opsplitter vi til to summer hvor den ene ikke afhænger af  $y_i$ .

$$\ln(L) = \sum_{i=1}^n \ln(1 - p(x_i)) + \sum_{i=1}^n y_i \cdot \ln\left(\frac{p(x_i)}{1 - p(x_i)}\right).$$

### Maksimum Likelihood med valgt $p(x)$

Nu har vi fået styr over udtrykket for likelihood, og vi har fundet et udtryk for  $p(x)$  ved den valgte model. Dog afhænger  $p(x)$  stadigvæk af valget af  $a$  og  $b$  hvorved  $L$  også kommer til at afhænge deraf, og vil derfor noteres som  $L(a, b)$  i det følgende.

Ved at kombinere ligningerne ??? får vi at

$$\ln(L(a, b)) = \sum_{i=1}^n \ln\left(1 - \frac{e^{a \cdot x_i + b}}{1 + e^{a \cdot x_i + b}}\right) + \sum_{i=1}^n y_i \cdot (a x_i + b).$$

Udtrykket i logaritmen laves om til en samlet brøk

$$\ln(L(a, b)) = \sum_{i=1}^n \ln\left(\frac{1}{1 + e^{a \cdot x_i + b}}\right) + \sum_{i=1}^n y_i \cdot (a x_i + b).$$

Her benytter vi igen regnereglen  $\ln(a - b) = \ln(a) - \ln(b)$ .

$$\ln(L(a, b)) = \sum_{i=1}^n \ln(1) - \ln(1 + e^{a \cdot x_i + b}) + \sum_{i=1}^n y_i \cdot (a x_i + b).$$

Da  $\ln(1) = 0$  har vi endelig

$$\ln(L(a, b)) = \sum_{i=1}^n -\ln(1 + e^{a \cdot x_i + b}) + \sum_{i=1}^n y_i \cdot (a x_i + b).$$

## Optimering

For at finde et minimum for  $L(a, b)$  eller  $\ln(L(a, b))$  kan vi selfølgelige finde stationære punkter ved at sætte de partiel afledte lig 0.

Lad os først se på  $\frac{\partial l(a, b)}{\partial b}$ . Ved den første skal vi se udtrykket som en sammensat funktion hvor den indre funktion har et led som også er en sammensat funktion.

$$\frac{\partial l(a, b)}{\partial b} = \sum_{i=1}^n -\frac{1}{1 + e^{a \cdot x_i + b}} \cdot (0 + e^{a \cdot x_i + b}) \cdot (0 \cdot x_i + 1) + \sum_{i=1}^n y_i \cdot (1 \cdot 0 + 1)$$

Ved at reducere fås

$$\frac{\partial l(a, b)}{\partial b} = \sum_{i=1}^n -\frac{e^{a \cdot x_i + b}}{1 + e^{a \cdot x_i + b}} + \sum_{i=1}^n y_i$$

Ved at bruge ligning ?? i forbindelse med den første sum kan dette omskrives til

$$\frac{\partial l(a, b)}{\partial b} = \sum_{i=1}^n -p(x_i) + \sum_{i=1}^n y_i = \sum_{i=1}^n y_i - p(x_i).$$

Nu ser vi på  $\frac{\partial l(a, b)}{\partial a}$ . på tilsvarende måde.

$$\frac{\partial l(a, b)}{\partial a} = \sum_{i=1}^n -\frac{1}{1 + e^{a \cdot x_i + b}} \cdot (0 + e^{a \cdot x_i + b}) \cdot (1 \cdot x_i + 0) + \sum_{i=1}^n y_i \cdot (1 \cdot x_i + 0)$$

Der reduceres

$$\frac{\partial l(a, b)}{\partial a} = \sum_{i=1}^n -\frac{e^{a \cdot x_i + b}}{1 + e^{a \cdot x_i + b}} \cdot x_i + \sum_{i=1}^n y_i \cdot x_i.$$

Igen bruges ligning ?? til at få

$$\frac{\partial l(a, b)}{\partial a} = \sum_{i=1}^n -p(x_i) \cdot x_i + \sum_{i=1}^n y_i \cdot x_i = \sum_{i=0}^n y_i \cdot x_i - p(x_i) \cdot x_i.$$



Endelig kan  $x_i$  sættes udenfor parentes hvorved vi har

$$\frac{\partial l(a, b)}{\partial b} = \sum_{i=0}^n (y_i - p(x_i)) \cdot x_i.$$

For at lave optimering og finde minimum skal vi derved løse følgende to ligninger med to ubekendte

$$0 = \sum_{i=0}^n (y_i - p(x_i)) \cdot x_i \quad \text{og} \quad 0 = \sum_{i=1}^n y_i - p(x_i).$$

Dette ligningssystem er dog ikke bare lige til at løse, så enten skal man igang med at benytte numeriske metoder til løsninger af dette eller benytte sig af gradient nedstigning (reference) for at få det optimale bund på parametrene  $a$  og  $b$  som fastlægger funktionen  $p(x)$ .

## Prediktion og Relation til logistisk udvikling

Når man endelig har fået fastlagt parametrene  $a$  og  $b$  kan vi endelig få vores model  $p(x)$  ud fra.

(mabgler alt med prædiktion og om man skal forudsige  $y = 1$  eller  $y = 0$ )

Man kan selvfølgelig tænke over om ovenstående skulle have relation til logistisk vækst. Hvis man fra logistisk klassifikation kommer frem til funktionen  $p(x)$

$$p(x) = \frac{1}{1 + e^{-a \cdot x - b}}.$$

Vi husker at det var denne funktionen som forudsiger sandsynligheden for  $y = 1$ .

Så f.eks. ved eksemplet med imun/ikke imun kan man forudsige hvad sandsynligheden er for at en tilfældig person er imun efter 100 dage ved at beregne

$$p(100) = \frac{1}{1 + e^{-a \cdot 100 - b}}.$$

Det kan selvfølgelig være at man ikke er interesseret i en person men istedet ønsker et bud på hvor mange der faktisk er immune

blandt hele befolkningen med  $M$  personer. Hvis  $f(x)$  er antal immune efter  $x$  dage må det være oplagt at benytte modellen

$$f(x) = M \cdot p(x) = \frac{M}{1 + e^{-a \cdot x - b}}.$$

Ved  $a$  kan vi lige gange og dividere med  $M$  og kalde  $a/M$  for  $a_{ny}$ . Derved er

$$f(x) = M \cdot p(x) = \frac{M}{1 + e^{-M \cdot a_{ny} \cdot x - b}} = \frac{M}{1 + e^{-M \cdot a_{ny} \cdot x - b}}.$$

Ved at benytte potensregneregler får vi

$$f(x) = M \cdot p(x) = \frac{M}{1 + e^{-M \cdot a_{ny} \cdot x - b}} = \frac{M}{1 + e^{-b} \cdot e^{-M \cdot a_{ny} \cdot x}}.$$

Hvis vi kalder  $e^{-b}$  for  $c$  får vi

$$f(x) = M \cdot p(x) = \frac{M}{1 + e^{-M \cdot a_{ny} \cdot x - b}} = \frac{M}{1 + c \cdot e^{-M \cdot a_{ny} \cdot x}}.$$

Som netop svarer til den kendte formel hørende til logistisk vækst og løsningen af differentiaalligningen

$$y' = a \cdot y \cdot (M - y)$$

.