Perceptroner

Aalborg Intelligence

Rosenblatts perceptron

Perceptroner... Hvorfor nu det? Jo, for det er faktisk sådan nogle, du ikke vidste, at du ikke kunne leve uden! Nu skal du høre hvorfor.

Eksempel: Hvem skal jeg stemme på ved næste valg?

De senere år er det blevet populært, at diverse medier laver forskellige kandidattests. Sådan nogle tests kan laves på mange forskellige måder - man kunne blandt andet bruge perceptroner! Testene fungerer som regel på den måde, at man bliver stillet en række forskellige spørgsmål og så skal man svare på en skala fra *meget uenig* til *meget enig*. Disse kategorier af svar kunne f.eks. oversættes til matematik på denne måde:

Helt enig	Overvejende enig	Hverken/eller	Overvejende uenig	Helt uenig
2	1	0	-1	-2

Lad os prøve at gøre det helt simpelt. I stedet for at komme med et bud på hvem man skal stemme på, så vil vi blot forsøge at komme med et bud på, om man skal stemme på rød eller blå blok (det er sikkert en håbløs simplificering, men det må du tale med din samfundsfagslærer om).

Lad os sige at vi vil basere vores bud på to spørgsmål:

1. Jeg synes, at indkomstskatten skal sættes ned.
2. Jeg synes ikke, at danske virksomheder skal pålægges en CO2-afgift.

Vi kan sikkert hurtigt blive enige om, at hvis man er meget enig i begge spørgsmål, så hører man formentlig til i blå blok og modsat, hvis man er meget uenig i begge spørgsmål, så hører man nok mere hjemme i rød blok. Så at lave en perceptron, som kan hjælpe os med at forudsige det, er nok ikke raketvidenskab, men det kan ikke desto mindre hjælpe os med at forstå de bagvedliggende principper og hvordan disse sidenhen kan generaliseres.

Lad os prøve at blive lidt mere specifikke og indføre to variable x_1 og x_2, hvor

- x_1: svaret på *Jeg synes, at indkomstskatten skal sættes ned* angivet på en skala fra -2 til 2
- x_2: svaret på *Jeg synes ikke, at danske virksomheder skal pålægges en CO2-afgift* angivet på en skala fra -2 til 2.

Vores beslutning vil vi nu også kvantificere vha. en variabel t, som kan antage to værdier, nemlig -1 og 1. Hvis vi hører hjemme i blå blok, vil vi sætte $t = 1$, mens vi vil sætte $t = -1$, hvis vi vil sætte vores krydset ved et rødt parti. Altså:

$$\begin{aligned} t = -1: &\quad \text{Rød blok} \\ t = 1: &\quad \text{Blå blok} \end{aligned}$$

Nu forestiller vi os, at vi har bedt seks personer (som godt ved, hvem de vil stemme på - måske er det ligefrem politikere vi har spurgt) om at svare på de to spørgsmål og samtidig tilkendegive, om de vil stemme på blå eller rød blok. Lad os f.eks. sige, at den første person er meget enig i at indkomstskatten skal sættes ned (dvs. $x_1 = 2$), og at denne person er overvejende enig i at danske virksomheder ikke skal pålægges en CO2-afgift (dvs. $x_2 = 1$). Desuden oplyser denne person, at han/hun vil stemme på blå blok (dvs. $t = 1$). Det kan udtrykkes sådan her:

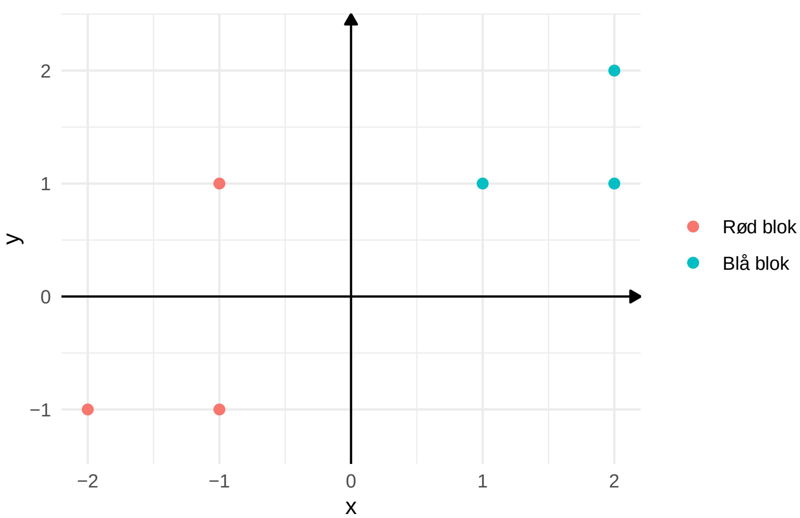
$$(x_1, x_2) = (2, 1) \quad \Rightarrow \quad t = 1 \tag{1}$$

Og sådan kunne man opstille andre eksempler:

$$
\begin{aligned}
(x_1, x_2) = (-1, 1) \quad &\Rightarrow \quad t = -1 \\
(x_1, x_2) = (-1, -1) \quad &\Rightarrow \quad t = -1 \\
(x_1, x_2) = (1, 1) \quad &\Rightarrow \quad t = 1 \\
(x_1, x_2) = (2, 2) \quad &\Rightarrow \quad t = 1 \\
(x_1, x_2) = (-2, -1) \quad &\Rightarrow \quad t = -1
\end{aligned}
$$

Det første eksempel siger for eksempel, at en person har været overvejende uenig i at sætte indkomstskatten ned ($x_1 = -1$), overvejende enig i at danske virksomheder ikke skal pålægges en CO2-afgift ($x_2 = 1$) og samtidig vil denne person stemme på rød blok ($t = -1$).

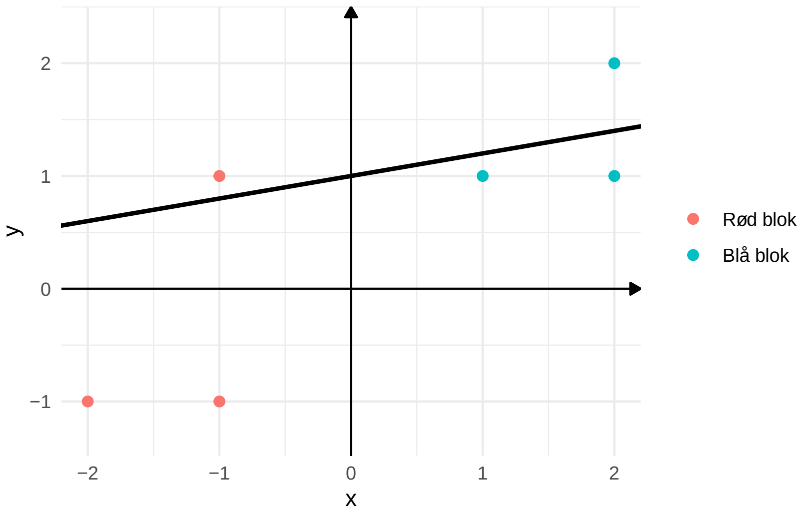
Vi kan prøve at indtegne (x_1, x_2)-punkterne i et koordinatsystem og samtidig angive den tilhørende værdi af t med en farve. Det vil se sådan her ud:



Figur 1: Illustration af svaret på spørgsmål 1 (1. aksen) og spørgsmål 2 (2. aksen) med en markering af om man vil stemme på rød eller blå blok.

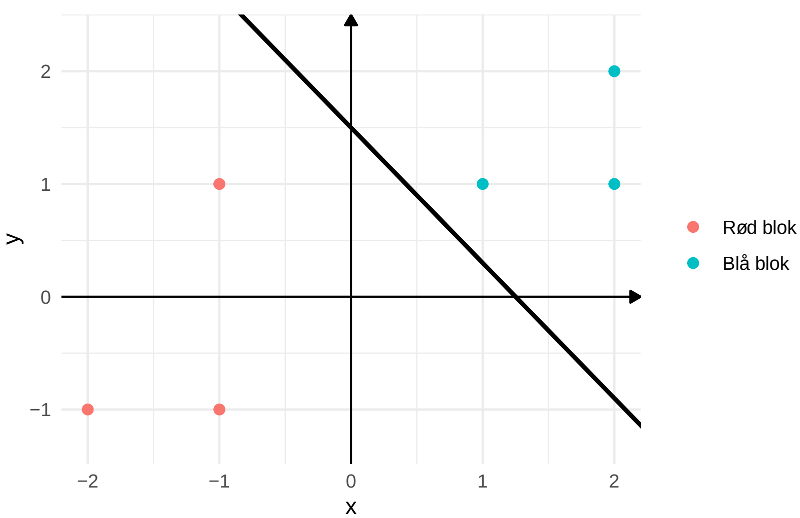
Det kunne godt se ud som om, at det vil være muligt at indtegne en ret linje på en sådan måde, at alle punkter som ligger over linjen skulle farves blå (svarende til "her stemmer vi på blå blok"), mens alle punkter under linjen skulle farves røde (svarende til "her stemmer vi på rød blok"). En tilfældig indtegnet linje ses på figur **??**.

Herunder ser du et bud på en linje, som ser ud til at være god til at adskille de blå punkter fra de røde – faktisk er der jo



Figur 2: Illustration af svaret på spørgsmål 1 (1. aksen) og spørgsmål 2 (2. aksen) med en markering af om man vil stemme på rød eller blå blok En tilfældig linje er indtegnet.

uendeligt mange linjer, som vil kunne adskille de blå punkter fra de røde:



Figur 3: Illustration af svaret på spørgsmål 1 (1. aksen) og spørgsmål 2 (2. aksen) med en markering af om man vil stemme på rød eller blå blok Her er indtegnet en linje, som kan separere de blå punkter fra de røde.

Linjen på figur **??** har ligning

$$y = -1.2 \cdot x + 1.5.$$

Men nu kaldte vi jo faktisk ikke de to variable for x og y, men derimod for x_1 og x_2. Med denne notation får vi altså, at

$$x_2 = -1.2 \cdot x_1 + 1.5$$

Hvis vi bruger denne ligning til at skelne imellem blå og røde punkter, så vil vi sige, at alle punkter, som ligger over linjen skal være blå. Det vil være det samme som at sige, at alle de blå punkter opfylder uligheden

$$x_2 > -1.2 \cdot x_1 + 1.5.$$

Eller skrevet på en anden måde:

$$1.2 \cdot x_1 + 1 \cdot x_2 > 1.5.$$

Her kalder man værdi 1.5 på højreside for *threshold* værdien (på dansk: tærskelværdi), fordi det er denne værdi, som afgør, om vi skal farve et punkt rødt eller blåt. Værdierne 1.2 og 1 kaldes for vægte, fordi de bestemmer, hvor meget inputværdierne x_1 og x_2 skal vægtes i forhold til hinanden.

En helt tredje måde at skrive det samme på vil være

$$-1.5 + 1.2 \cdot x_1 + 1 \cdot x_2 > 0.$$

Nu kalder man så bare værdien -1.5 for en bias, men i virkeligheden er det jo bare threshold værdien med modsat fortegn[1].

[1] Der er forskellige overvejelser i forhold til valget af denne skrivemåde. For det første er vi gået væk fra x og y og over til x_1 og x_2. Det giver mening, fordi vi ofte tænker på y som den afhængige variabel og x som den uafhængige variabel. Denne fortolkning af de to variable giver ikke mening i denne sammenhæng. Derudover kan vi beskrive en vilkårlig linje i planen ved hjælp af ligningen $ax_1 + bx_2 + c = 0$ – også de lodrette linjer. Holder vi derimod fast i $y = ax+b$, så kan vi ikke "fange" de lodrette linjer.

Vi har nu faktisk udledt en regel, som for tid og evighed kan hjælpe os med at afgøre, om vi skal stemme på rød eller blå blok. Den kan opsummeres sådan her:

Hvem skal jeg stemme på?

Svar på en skala fra -2 til 2 på følgende spørgsmål:
x_1: "Jeg synes, at indkomstskatten skal sættes ned"
x_2: "Jeg synes ikke, at danske virksomheder skal pålægges en CO2-afgift"
hvor 2 svarer til "Meget enig" og -2 svarer til "Meget uenig".
Beregn nu o (for outputværdi) på denne måde

$$o = \begin{cases} 1 & \text{hvis } -1.5 + 1.2 \cdot x_1 + 1 \cdot x_2 \geq 0 \\ -1 & \text{hvis } -1.5 + 1.2 \cdot x_1 + 1 \cdot x_2 < 0. \end{cases}$$

Reglen er nu:

$$\begin{aligned} \text{Hvis } o = 1: & \quad \text{Stem blå blok.} \\ \text{Hvis } o = -1: & \quad \text{Stem rød blok.} \end{aligned}$$

Man siger også, at man på baggrund af inputværdierne kan lave en klassificering (eller kategorisering). Det betyder, at vi på baggrund af inputværdierne kan beregne, om vi er i kategorien ”Blå blok” ($o = 1$) eller i kategorien ”Rød blok” ($o = -1$). Grafisk svarer det til, at man indtegner sit (x_1, x_2)-punkt i koordinatsystemet i figur **??** og ser så på om punkt ligger over eller under linjen (ligger det over skal vi stemme blå blok).

Eksempel 0.1. *Lad os sige at en vælger hverken er enig eller uenig i, at indkomstskatten skal sættes ned. Det vil sige, at* $x_1 = 0$. *Samtidig er denne vælger meget enig i, at danske virksomheder ikke skal pålægges en CO2-afgift. Altså er* $x_2 = 2$. *Vi udregner nu:*

$$-1.5 + 1.2 \cdot x_1 + x_2 = -1.5 + 1.2 \cdot 0 + 2 = 0.5$$

Og da denne værdi er større end 0*, sætter vi* $o = 1$*. Det vil sige, at vi vil anbefale denne vælger at stemme blå blok.*

Det er da smart! Og det her er faktisk lige præcis idéen bag perceptroner, som den amerikanske psykolog Frank Rosenblatt foreslog helt tilbage i 1958. Den klassiske perceptron er defineret ved, at perceptronen kan modtage input

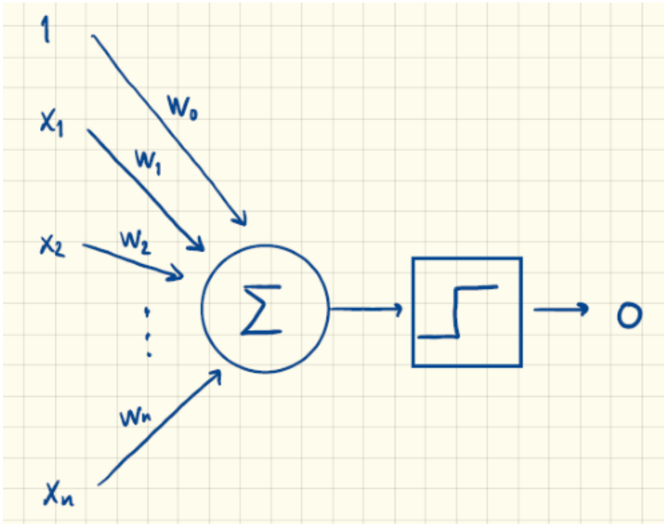
$$x_1, x_2, \dots, x_n,$$

hvor hver enkel inputværdi i princippet kan være et vilkårligt reelt tal. I vores eksempel har vi dog begrænset inputværdierne til $x_1, x_2 \in \{-2, -1, 0, 1, 2\}$. Vi beregner så en outputværdi o vha. vægtene $w_1, w_2, \dots, w_n$ og en biasværdi, som vi her vil kalde for w_0 på denne måde:

$$o = \begin{cases} 1 & \text{hvis } w_0 + w_1 \cdot x_1 + \cdots + w_n \cdot x_n \geq 0 \\ -1 & \text{hvis } w_0 + w_1 \cdot x_1 + \cdots + w_n \cdot x_n < 0. \end{cases}$$

Grafisk kan det illustreres sådan her:

Her illustrerer sumtegnet i cirklen, at vi tager en vægtet sum af alle inputværdierne (inklusiv et input (x_0), som altid er 1, og som vægtes med w_0 svarende til, at vi får vores bias med), mens grafen af trappefunktionen i firkanten viser, at vi diskretiserer denne vægtede sum, sådan at outputværdien enten er -1 eller 1.



Figur 4: Grafisk illustration af en perceptron.

VIDEO: Hvad er en perceptron?

I denne video forklarer vi ovenstående, men med udgangspunkt i et andet eksempel.

https://www.youtube.com/embed/1WTZZCx-pRY

Perceptron Learning Algoritmen

Det er jo alt sammen meget fint, hvis man kender vægtene. Bum – så kan man beregne den vægtede sum

$$w_0 + w_1 \cdot x_1 + \cdots + w_n \cdot x_n.$$

Får vi et ikke-negativt tal sættes o til 1 og -1 ellers. Præcis som vi gjorde det i eksempel **??**. Men det kræver jo, at man kender vægtene $w_0, w_1, \dots, w_n$...! I vores eksempel med bare to inputværdier var det nemt nok at finde nogle passende værdier af vægtene. Vi tegnede bare punkterne (x_1, x_2) ind i et koordinatsystem og slog en streg, der adskilte de røde punkter fra de blå punkter. Men hvis der er mere end to inputværdier (hvis man f.eks. i kandidattesten skal svare på 20 spørgsmål), så er det jo ikke helt så nemt! Hvad gør vi så?

Det havde Frank Rosenblatt faktisk også en idé til, som vi nu skal se nærmere på. Vi skal forestille os, at vi har en masse data, som vi så det i vores eksempel. Disse data kalder vi for træningsdata, og de vil dels bestå af konkrete inputværdier $x_1, x_2, \ldots, x_n$ og den tilhørende korrekte klassificering t. Her bruger vi bogstavet t som en forkortelse for *target-værdi* – altså den "målværdi", som svarer til den rigtige klassificering. I vores eksempel viser (**??**) et eksempel på sådan et træningsdatasæt.

Lad os så se på hvad Rosenblatts "Perceptron Learning Algoritme" går nu ud på. Den kommer her:

- Sæt alle vægte $w_0, w_1, \ldots, w_n$ til et tilfældigt tal (f.eks. 0.5).
- Tag et træningseksempel $(x_1, x_2, \ldots, x_n)$ med tilhørende target-værdi t.
- Udregn outputværdien o:

$$o = \begin{cases} 1 & \text{hvis } w_0 + w_1 \cdot x_1 + \cdots + w_n \cdot x_n \geq 0 \\ -1 & \text{hvis } w_0 + w_1 \cdot x_1 + \cdots + w_n \cdot x_n < 0. \end{cases}$$

- Udregn fejlen *error* (som er forskellen mellem den ønskede target-værdi t og den beregnede outputværdi o):

$$error = t - o$$

- Opdatér alle vægtene:

$$\begin{aligned} w_0 &\leftarrow w_0 + \eta \cdot error \\ w_1 &\leftarrow w_1 + \eta \cdot error \cdot x_1 \\ w_2 &\leftarrow w_2 + \eta \cdot error \cdot x_2 \\ &\vdots \\ w_n &\leftarrow w_n + \eta \cdot error \cdot x_n \end{aligned}$$

  Her er η (udtales "eta") et tal mellem 0 og 1, som kaldes for en *learning rate* (på dansk: en læringsrate). Værdien af η afgør hvor hurtigt, vi skal opdatere vægtene. Hvis η er tæt på 0 opdateres vægtene langsomt, hvorimod en værdi af η tæt på 1 vil betyde, at vægtene opdateres hurtigt.