Proefstudeerdag Wiskunde: Hoe computerprogramma's kunnen leren

Werkcollege, maandag 18 mei

We gaan laten zien dat de functie $F(w)$ uit het college convex is, voor het geval van één parameter w. Het gaat dan om de volgende functie:

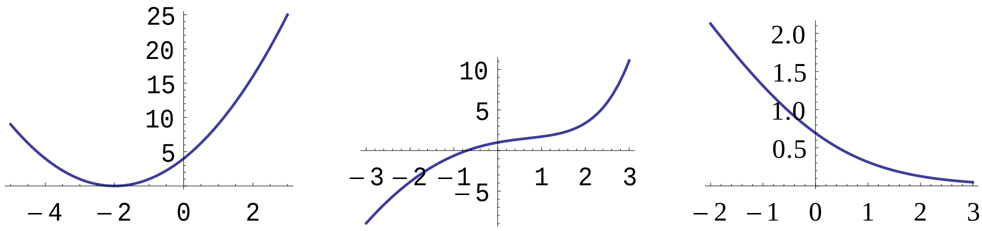
$$F(w) = \ln(1 + e^{-wx_1y_1}) + \ln(1 + e^{-wx_2y_2}) + \ldots + \ln(1 + e^{-wx_{60\,000}y_{60\,000}}), \qquad (\star)$$

waarbij $x_1, x_2, \ldots, x_{60\,000}$ getallen zijn waar we niets over aannemen, en van de getallen $y_1, y_2, \ldots, y_{60\,000}$ weten we weliswaar dat ze allemaal -1 of $+1$ zijn, maar dat zullen we niet hoeven te gebruiken.

De definitie van een convexe functie f is dat voor iedere twee punten w en v geldt dat

$$f\big((1-\lambda)w + \lambda v\big) \leq (1-\lambda)f(w) + \lambda f(v) \qquad \text{voor alle } 0 \leq \lambda \leq 1.$$

Intuïtief betekent dit dat als we een rechte lijn tussen twee punten van de functie tekenen, dat de functie daar dan altijd zelf onder blijft. We kunnen ons nu afvragen of de functies in Figuur 1 hieronder convex zijn.



(a) $f(w) = (w+2)^2$ (b) $f(w) = e^w - w^2$ (c) $f(w) = \ln(1 + e^{-w})$

Figuur 1: Drie al dan niet convexe functies

Opgave 1. Zoals je kunt zien in de grafiek is de functie $f(w) = e^w - w^2$ niet convex in de buurt van $w = 0$. Laat zien dat inderdaad niet aan de definitie wordt voldaan als we $w = -1$, $v = 1$ en $\lambda = 1/2$ invullen. Hint: $e \approx 2,7$ en $e^{-1} \approx 0,4$.
ANTWOORD: De definitie van convexiteit vereist dat

$$f\Big((1-\lambda)w + \lambda v\Big) = f\Big(\tfrac{1}{2} \times (-1) + \tfrac{1}{2} \times 1\Big) = f(0) = 1$$

niet meer is dan

$$(1-\lambda)f(w) + \lambda f(v) = \tfrac{1}{2}f(-1) + \tfrac{1}{2}f(1) = \tfrac{1}{2}(e^{-1} - 1) + \tfrac{1}{2}(e - 1) = 0,543\ldots$$

Maar daar wordt niet aan voldaan.

De andere twee functies lijken wel convex te zijn, maar hoe kunnen we dat nou echt zeker weten? Om deze vraag te beantwoorden, is het handig om wat wiskundig gereedschap op te bouwen. Ons eerste stuk gereedschap is het volgende:

Stelling 1. Als $f(w)$ een convexe functie is, dan is $f(aw + b)$ dat ook voor arbitraire getallen a and b.

Opgave 2. Hieronder zullen we aantonen dat de functie $f(w) = w^2$ convex is. Gebruik Stelling 1 om te laten zien dat $f(w) = (w+2)^2$ dan ook convex moet zijn.
ANTWOORD: Pas Stelling 1 toe met $a = 1$ en $b = 2$.

Opgave 3. Kun je Stelling 1 bewijzen, gebruikmakend van de definitie van convexiteit?
ANTWOORD: We moeten laten zien dat $g(w) = f(aw + b)$ convex is, gebruikmakend van het

feit dat $f(w)$ convex is. Laat w en v twee arbitraire punten zijn, en kies ook voor $\lambda \in [0,1]$ wat je wilt. Definieer vervolgens $x = aw + b$ en $y = av + b$. Dan geldt:

$$
\begin{aligned}
g\Big((1-\lambda)w + \lambda v\Big) = f\Big(a\big((1-\lambda)w + \lambda v\big) + b\Big) &= f\Big((1-\lambda)x + \lambda y\Big) \\
&\le (1-\lambda)f(x) + \lambda f(y) = (1-\lambda)g(w) + \lambda g(v),
\end{aligned}
$$

en dat is wat we moesten laten zien. Hierbij hebben we voor de ongelijkheid gebruik gemaakt van de convexiteit van f.

Het is mogelijk om de definitie van convexiteit direct te controleren voor $f(w) = w^2$, maar dit vergt de nodige moeite. Veel makkelijker wordt het met behulp van de volgende stelling, die tijdens de studie wiskunde bewezen wordt:

Stelling 2. Als de tweede afgeleide van een functie f altijd groter of gelijk aan 0 is, dan is f een convexe functie.

Opgave 4. Gebruik Stelling 2 om aan te tonen dat $f(w) = w^2$ en $f(w) = \ln(1 + e^w)$ beide convexe functies zijn.
ANTWOORD: De tweede afgeleide van $f(w) = w^2$ is $f''(w) = 2 > 0$. En de eerste afgeleide f' en tweede afgeleide f'' van $f(w) = \ln(1 + e^w)$ zijn:

$$
\begin{aligned}
f'(w) &= \frac{e^w}{1+e^w}, \\
f''(w) &= \frac{e^w}{1+e^w} - \frac{e^w}{(1+e^w)^2} = \frac{e^w(1+e^w) - e^w}{(1+e^w)^2} = \frac{e^w}{(1+e^w)^2} > 0.
\end{aligned}
$$

Dus beide afgeleides zijn altijd positief, en het volgt uit Stelling 2 dat beide functies convex zijn.

Opgave 5. Laat nu met behulp van Stelling 1 zien dat $f(w) = \ln(1 + e^{-w})$ ook een convexe functie is.
ANTWOORD: We weten uit de vorige opgave dat $\ln(1 + e^w)$ een convexe functie is. Pas daarop Stelling 1 toe met $a = -1$ en $b = 0$.

We willen zo direct aantonen dat de functie F uit het college een convexe functie is. Daarvoor maken we het onszelf vele malen makkelijker als we het volgende weten:

Stelling 3. Als $f(w)$ en $g(w)$ beide convexe functies zijn, dan is $f(w) + g(w)$ dat ook.

Opgave 6. Kun je op basis van de definitie van convexe functies Stelling 3 bewijzen?
ANTWOORD: Laat f en g arbitraire convexe functies zijn. Dan geldt voor iedere punten w en v en iedere keuze van $\lambda \in [0,1]$ dat:

$$
\begin{aligned}
f\Big((1-\lambda)w + \lambda v\Big) + g\Big((1-\lambda)w + \lambda v\Big) &\le (1-\lambda)f(w) + \lambda f(v) + (1-\lambda)g(w) + \lambda g(v) \\
&= (1-\lambda)\Big(f(w) + g(w)\Big) + \lambda\Big(f(v) + g(v)\Big),
\end{aligned}
$$

waarmee we hebben laten zien dat $f(w) + g(w)$ ook convex is.

Opgave 7. Gebruik de benodigde stellingen om te laten zien dat

$$
F_2(w) = \ln(1 + e^{-wx_1y_1}) + \ln(1 + e^{-wx_2y_2})
$$

een convexe functie is voor alle mogelijke getallen x_1, y_1, x_2, y_2.
ANTWOORD: We weten uit Opgave 5 dat $\ln(1 + e^{-w})$ een convexe functie is. Het volgt vervolgens uit Stelling 1 met $a = x_1y_1$ of $a = x_2y_2$ en $b = 0$ dat $\ln(1 + e^{-wx_1y_1})$ en $\ln(1 + e^{-wx_2y_2})$ beide convex zijn. Pas vervolgens Stelling 3 toe om te laten zien dat hun som $F_2(w)$ ook convex is.

Opgave 8. We zijn er nu bijna, behalve dat de functie $F(w)$ uit $(\star)$ niet bestaat uit een som van twee functies, maar uit een som van 60 000 functies. Kun je een manier verzinnen om het bewijs compleet te maken dat de functie uit $(\star)$ convex is?
ANTWOORD: Deze vraag zat er in als moeilijke uitsmijter. Het bewijs gaat als volgt. Laat

$$
F_n(w) = \ln(1 + e^{-wx_1y_1}) + \ldots + \ln(1 + e^{-wx_ny_n})
$$

de som zijn van de eerste n functies. We willen bewijzen dat F_n convex is voor $n = 60\,000$, maar daar zijn we nog niet. Om daar te komen is het voldoende om de volgende twee punten te bewijzen:

1. $F_1(w)$ is convex.
2. Stel dat $F_n(w)$ convex is, dan is $F_{n+1}(w)$ dat ook.

Als we de twee regels combineren voor $n = 1$, dan weten we dat $F_2(w)$ convex is, en als we dat dan nog een keer combineren met regel 2, dan weten we dat $F_3(w)$ ook convex is, en zo door tot we weten dat $F_{60\,000}$ convex is. Dit type argument wordt een bewijs met volledige inductie genoemd.

Hoe bewijzen we nu die twee punten? Om te beginnen kunnen we observeren dat $\ln(1 + e^{-wx_iy_i})$ convex is voor iedere $i = 1, \ldots, 60\,000$, want dat volgt weer uit Stelling 1 met $a = x_iy_i$ en $b = 0$. In het bijzonder weten we dus dat $F_1(w)$ convex is. Vervolgens observeren we dat

$$F_{n+1}(w) = F_n(w) + \ln(1 + e^{-wx_{n+1}y_{n+1}}),$$

dus als $F_n(w)$ convex is, dan volgt het uit Stelling 3 dat $F_{n+1}(w)$ dat ook moet zijn, en daarmee zijn we klaar.