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:

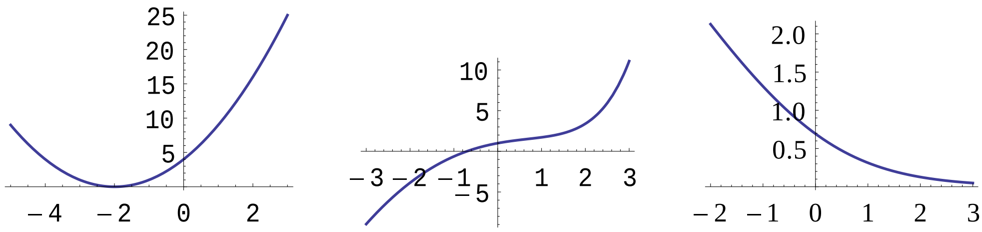
$$F(w) = \ln(1 + e^{-wx_1y_1}) + \ln(1 + e^{-wx_2y_2}) + \ldots + \ln(1 + e^{-wx_{60\,000}y_{60\,000}}), \qquad (\star)$$

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 $e^{-1} \approx 0.4$.

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.

Opgave 3. Kun je Stelling 1 bewijzen, gebruikmakend van de definitie van convexiteit?

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.

Opgave 5. Laat nu met behulp van Stelling 1 zien dat $f(w) = \ln(1 + e^{-w})$ ook een convexe functie is.

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?

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

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

een convexe functie is voor alle mogelijke getallen x_1, y_1, x_2, y_2.

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?