Het 2-3-1 XOR-Netwerk heeft lokale Minima[†]

Ida G. Sprinkhuizen-Kuyper
Egbert J.W. Boers
Vakgroep Informatica
RijksUniversiteit Leiden
Postbus 9512
2300 RA Leiden
{kuyper,boers}@wi.leidenuniv.nl

Samenvatting

In dit artikel wordt een lokaal (niet absoluut) minimum gegeven voor een tweelagig netwerk met drie verborgen knopen dat de vier patronen van de XOR-functie leert. Dit is een weerlegging van artikelen van Poston e.a. [2] en Yu [6] waarin bewezen wordt dat tweelagige neurale netwerken met t-1 verborgen knopen, die een trainingsverzameling van t voorbeelden leren, geen lokale minima kunnen hebben.

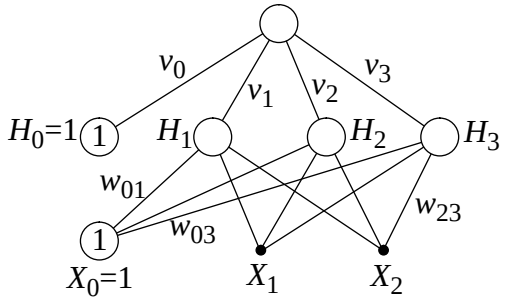
1 Inleiding

Recentelijk hebben wij de foutoppervlakken van de twee eenvoudigste neurale netwerken voor de XOR-functie bestudeerd [3, 4, 5]. Er waren wel enige resultaten over deze foutoppervlakken, maar er was nog geen volledige classificatie van alle stationaire punten. Ons idee is dat een onderzoek van de stationaire punten van foutoppervlakken van simpele neurale netwerken inzicht kan geven in de problemen die een leeralgoritme tegen kan komen en waar een leeralgoritme een oplossing voor zou moeten hebben. Bij het onderzoek van het 2-2-1 XOR-netwerk (2 invoereenheden, een verborgen laag met 2 knopen en een uitvoerlaag met 1 knoop) vonden wij een aantal lokale minima, die met elkaar gemeen hadden dat één of twee verborgen knopen verzadigd waren voor twee of meer van de te leren patronen [4, 5].

Van sommige van deze lokale minima was het voor de hand liggend dat het toevoegen van een derde (vierde enz.) knoop zou leiden tot een analoog lokaal minimum. Dit betekent dus dat we een probleem hebben met vier patronen, waarvoor een tweelagig neuraal netwerk met drie of meer verborgen knopen wel degelijk lokale minima heeft.

Dit resultaat is in tegenspraak met de resultaten van Poston e.a. [2] en Yu [6]. In het artikel van Poston e.a. wordt bewezen dat voor “normale” problemen” (dat wil zeggen geen tegenstrijdige invoer) met t trainingsvoorbeelden geldt dat een neuraal netwerk met t verborgen knopen (inclusief een verborgen knoop met vaste uitvoer 1 voor de drempelwaarde van de uitvoerknoop, dus $t-1$ “echte” verborgen knopen) deze trainingsvoorbeelden exact kan representeren met fout nul. Ook wordt bewezen dat zo’n neuraal

[†]. Technical Report 96-19, Dept. of Computer Science, Leiden University. Available as ftp://ftp.wi.leidenuniv.nl/pub/CS/TechnicalReports/1996/tr96-19.ps.gz



Figuur 1: Het 2-3-1 XOR-netwerk

netwerk geen lokale minima (minima met een fout groter dan nul) kan hebben. Yu [6] generaliseert dit resultaat tot trainingsverzamelingen met tegenstrijdigheden. Daar zal geen fout gelijk aan nul gevonden kunnen worden, maar wel een absoluut minimum voor de fout. Het bewijs dat Yu geeft wordt bekritiseerd door Hamey [1]. In een reactie op de kritiek van Hamey corrigeert Yu zijn bewijs tot een bewijs dat in essentie ook door Poston gegeven is.

Het idee achter de gegeven bewijzen is, dat als bewezen kan worden dat vanuit bijna ieder punt (dat wil zeggen alle punten op een verzameling van maat nul na) uit de gewichtenruimte er een rechtlijnig pad is met strikt dalende fout naar een punt met fout nul, dat dan ook de overgebleven punten geen lokale minima kunnen zijn. Het argument is dat willekeurig dicht bij zo'n punt een punt ligt vanwaar een strikt dalend pad loopt, zodat zo'n punt niet gescheiden kan worden van een zogenaamd "goed" punt door een barrière van welke positieve hoogte dan ook.

Dit blijkt dus niet zo te zijn: er zijn gebieden met lokale minima, waarbij voor ieder punt uit zo'n gebied geldt dat in een omgeving alleen punten liggen met gelijke fout of wel grotere fout, terwijl vanuit ieder punt met een grotere fout er een rechte weg met strikt dalende fout is naar een punt met absoluut minimale foutwaarde.

In paragraaf 2 laten we zien dat het 2-3-1 XOR-netwerk een lokaal minimum heeft. In paragraaf 3 geven we een uitwerking van het bewijs van Poston voor het 2-3-1 XOR-netwerk en in paragraaf 4 geven we een aantal conclusies.

2 Het 2-3-1 XOR-netwerk en een lokaal minimum voor dit netwerk

Het 2-3-1 XOR-netwerk is gegeven in figuur 1. De invoer wordt gevormd door X_1 en X_2. Een extra invoer X_0 altijd gelijk aan 1 is toegevoegd voor de drempelwaarde. De gewichten naar de verborgen knopen worden aangeduid met w_{ij}, waarbij i het nummer van de betreffende invoer is ($i \in \{0, 1, 2\}$) en j het nummer van de verborgen knoop ($j \in \{1, 2, 3\}$). Voor de drempelwaarde van de uitvoerknoop is een extra verborgen knoop H_0 toegevoegd met constante uitvoer 1. De gewichten van de verborgen knopen naar de uitvoerknoop worden aangeduid met v_j, waarbij j het nummer van de verborgen knoop is ($j \in \{0, 1, 2, 3\}$).

We zullen voor de gewichten van de invoer naar de verborgen laag de notatie $\boldsymbol{w}$ gebruiken en de gewichten van de verborgen laag naar de uitvoerknoop zullen we aanduiden met $\boldsymbol{v}$. Voor alle gewichten van het netwerk gebruiken we de notatie $\boldsymbol{W}$. Dus

$\boldsymbol{w} = (w_{01}, w_{11}, w_{21}, w_{02}, w_{12}, w_{22}, w_{03}, w_{13}, w_{23})$,
$\boldsymbol{v} = (v_0, v_1, v_2, v_3)$ en
$\boldsymbol{W} = (\boldsymbol{w}, \boldsymbol{v}) = (w_{01}, w_{11}, w_{21}, w_{02}, w_{12}, w_{22}, w_{03}, w_{13}, w_{23}, v_0, v_1, v_2, v_3)$.
Dit netwerk moet de XOR-functie leren zoals gegeven in tabel 1.

Tabel 1: Patronen voor het XOR-probleem

Patroon	X_1	X_2	**gewenste uitvoer**
P_{00}	0	0	0.1
P_{01}	0	1	0.9
P_{10}	1	0	0.9
P_{11}	1	1	0.1

Iedere knoop gebruikt als overdrachtsfunctie de sigmoide $f(x) = 1/(1+e^{-x})$. De invoer van de uitvoerknoop voor patroon P_{ij} zullen we aangeven met A_{ij}, dus

$$A_{00} = v_0 + v_1 f(w_{01}) + v_2 f(w_{02}) + v_3 f(w_{03})$$
$$A_{01} = v_0 + v_1 f(w_{01} + w_{21}) + v_2 f(w_{02} + w_{22}) + v_3 f(w_{03} + w_{23})$$
$$A_{10} = v_0 + v_1 f(w_{01} + w_{11}) + v_2 f(w_{02} + w_{12}) + v_3 f(w_{03} + w_{13})$$
$$A_{11} = v_0 + v_1 f(w_{01} + w_{11} + w_{21}) + v_2 f(w_{02} + w_{12} + w_{22}) + v_3 f(w_{03} + w_{13} + w_{23})$$

Het netwerk heeft dus de 4 patronen van de XOR-functie geleerd als

$$f(A_{00}) = 0.1$$
$$f(A_{01}) = 0.9$$
$$f(A_{10}) = 0.9$$
$$f(A_{11}) = 0.1$$

Voor de fout gemaakt door het netwerk beschouwen we de kwadratische fout:

$$E = \frac{1}{2}(f(A_{00}) - 0.1)^2 + \frac{1}{2}(f(A_{01}) - 0.9)^2 + \frac{1}{2}(f(A_{10}) - 0.9)^2 + \frac{1}{2}(f(A_{11}) - 0.1)^2$$

We zullen de volgende stelling bewijzen, die zegt dat dit netwerk lokale minima heeft:

Stelling 2.1 *Het foutoppervlak van het 2-3-1 netwerk, zoals gegeven in figuur 1, voor het XOR-probleem heeft een lokaal minimum met* $E = 0.32$ *als aan de volgende voorwaarden is voldaan:*

$$w_{01} = w_{01} + w_{11} = w_{01} + w_{21} = w_{01} + w_{11} + w_{21} = \infty$$
$$w_{02} = w_{02} + w_{12} = w_{02} + w_{22} = w_{02} + w_{12} + w_{22} = \infty$$
$$w_{03} = w_{03} + w_{13} = w_{03} + w_{23} = w_{03} + w_{13} + w_{23} = \infty$$

(dus de verborgen knopen zijn voor alle 4 de patronen verzadigd)
$v_0 + v_1 + v_2 + v_3 = 0$
$v_1 w_{11} w_{21} < 0$
$v_2 w_{12} w_{22} < 0$
$v_3 w_{13} w_{23} < 0$

Bewijs In de gegeven punten geldt dat $A_{ij} = 0$ en dus $f(A_{ij}) = 0.5$ voor alle patronen P_{ij}. De fout E is dus 0.32. Om te zien wat er gebeurt als het gewicht w_{01} het oneindige verlaat, voeren we de hulpvariabele $p_1 = e^{-w_{01}}$ in. Differentiëren we de foutfunctie E partieel naar p_1, dan vinden we:

$$\left.\frac{\partial E}{\partial p_1}\right|_{p_1 = 0} = -(f(0) - 0.1)\, f'(0)\, v_1\, (1 - e^{-w_{11}})\, (1 - e^{-w_{21}})$$

Hieruit volgt dat als p_1 stijgt vanuit nul (dus w_{01} wordt kleiner dan oneindig), dat dan de fout zal stijgen als $v_1 w_{11} w_{21} < 0$. Net zo wordt aangetoond dat de fout zal stijgen als w_{02} en/of w_{03} kleiner dan oneindig worden en $v_2 w_{12} w_{22} < 0$ en $v_3 w_{13} w_{23} < 0$. De fout zal dus in deze gevallen alleen kleiner gemaakt kunnen worden door de gewichten v_0, v_1, v_2 en v_3 te veranderen, want zolang w_{0i} oneindig is, hebben de gewichten w_{1i} en w_{2i} geen invloed op de fout. De fout als functie van $x = v_0 + v_1 + v_2 + v_3$ is echter gelijk aan $(f(x) - 0.1)^2 + (f(x) - 0.9)^2$ en deze functie heeft een minimum voor $x = 0$. Het gevolg is dat voor alle punten in een omgeving van de genoemde punten geldt dat $E \geq 0.32$ en dus zijn de gegeven punten lokale minima. ❑

In de volgende paragraaf zullen we laten zien dat voor bijna alle[†] punten van het fout-oppervlak van dit neurale netwerk voor het XOR-probleem een recht pad bestaat met strikt dalende fout naar een punt met fout nul. Dit geldt dus ook voor punten willekeurig dicht bij de punten van stelling 2.1.

3 Het bewijs van Poston e.a. voor het 2-3-1 XOR-netwerk

Het bewijs van Poston e.a. berust op het beschouwen van de matrix U die gevormd wordt door de vectoren van de uitvoer van de verborgen knopen voor alle patronen. Voor het 2-3-1 netwerk voor het XOR-probleem wordt

$$U = \begin{bmatrix} 1 & f(w_{01}) & f(w_{02}) & f(w_{03}) \\ 1 & f(w_{01} + w_{21}) & f(w_{02} + w_{22}) & f(w_{03} + w_{23}) \\ 1 & f(w_{01} + w_{11}) & f(w_{02} + w_{12}) & f(w_{03} + w_{13}) \\ 1 & f(w_{01} + w_{11} + w_{21}) & f(w_{02} + w_{12} + w_{22}) & f(w_{03} + w_{13} + w_{23}) \end{bmatrix}$$

Als deze matrix een determinant ongelijk aan nul heeft, dan zijn er waarden te vinden van de gewichten van de verborgen knopen naar de uitvoerknoop zo, dat de fout gelijk aan nul

[†]. Als we spreken over “bijna alle” punten in een n-dimensionale ruimte, dan bedoelen we alle punten op een verzameling van maat nul na. Iedere variëteit van dimensie kleiner dan n heeft typisch maat nul. Een punt uit een verzameling van maat nul in een Euclidische ruimte heeft de eigenschap dat willekeurig dicht bij zo’n punt punten liggen die niet tot die verzameling behoren.

is, want dan is er een oplossing $\boldsymbol{v} = (v_0, v_1, v_2, v_3)$ te vinden van het stelsel vergelijkingen:

$$U \begin{bmatrix} v_0 \\ v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} f^{-1}(0.1) \\ f^{-1}(0.9) \\ f^{-1}(0.9) \\ f^{-1}(0.1) \end{bmatrix}$$

Omdat U een analytische functie is van de gewichten $\boldsymbol{w}$ van de invoer naar de verborgen knopen, geldt dat ofwel $det(U) = 0 \ \forall \boldsymbol{w}$, ofwel de verzameling van punten waar $det(U) = 0$ vormt een verzameling van maat nul in de ruimte van de gewichten $\boldsymbol{w}$. Het is eenvoudig na te gaan dat $det(U)$ niet gelijk is aan nul voor alle $\boldsymbol{w}$. Beschouwen we nu de ruimte van alle gewichten $\boldsymbol{W}$. Dan geldt dus voor bijna alle gewichten $\boldsymbol{W}$ dat $det(U) \neq 0$. Er geldt nu de volgende stelling:

Stelling 3.1 *Zij $\boldsymbol{W}$ een punt in de gewichtenruimte met $det(U) \neq 0$, dan is er een recht pad met strikt dalende fout van $\boldsymbol{W}$ naar een gewichtenvector $\tilde{\boldsymbol{W}}$, waarvoor de fout gelijk is aan nul.*

Bewijs Zij $\boldsymbol{W} = (\boldsymbol{w}, \boldsymbol{v})$ en zij $\tilde{\boldsymbol{W}} = (\boldsymbol{w}, \tilde{\boldsymbol{v}})$ met $\tilde{\boldsymbol{v}}$ de oplossing van het stelsel vergelijkingen $U\tilde{\boldsymbol{v}} = (f^{-1}(0.1), f^{-1}(0.9), f^{-1}(0.9), f^{-1}(0.1))^T$. Op de lijn van $\boldsymbol{W}$ naar $\tilde{\boldsymbol{W}}$, die gegeven wordt door $\boldsymbol{W}(\lambda) = \boldsymbol{W} + \lambda(\tilde{\boldsymbol{W}} - \boldsymbol{W}) \ (0 \leq \lambda \leq 1)$ geldt dat de invoer van de uitvoerknoop voor ieder patroon monotoon naar de gewenste invoer gaat. Omdat de overdrachtsfunctie f strikt monotoon stijgend is, zal op deze lijn ook gelden dat de fout voor ieder patroon strikt monotoon daalt en dus daalt ook de totale fout E op deze lijn. Als $\lambda = 1$ dan is het punt $\tilde{\boldsymbol{W}}$ bereikt en is de fout nul. ❑

Er geldt dus dat voor bijna alle gewichten $\boldsymbol{W}$ er een strikt dalend pad is naar een punt in de gewichtenruimte met fout nul.

In de vorige paragraaf hebben we gezien dat dit niet tot gevolg heeft dat er geen lokale minima kunnen zijn, hoewel willekeurig dicht bij zo'n lokaal minimum een punt te vinden is vanwaar een strikt dalend pad naar een absoluut minimum bestaat.

Dit betekent dat, hoewel er willekeurig dicht bij zo'n lokaal minimum een punt is, van waaruit er een rechtstreeks pad is naar een absoluut minimum, de fout van zo'n punt groter is dan die van het lokale minimum en het pad zo langzaam zal dalen, dat de conclusie, dat er willekeurig dicht bij het lokale minimum ook een punt met lagere fout moet liggen, niet gerechtvaardigd is.

4 Conclusie

In het voorafgaande hebben we laten zien dat in een neuraal netwerk met een verborgen laag met 3 knopen lokale minima kunnen optreden wanneer de vier patronen van het XOR-probleem geleerd worden. Dit is in tegenspraak met de resultaten van Poston e.a. [2] en Yu [6] (zie ook Hamey [1]). Het is mogelijk om vanuit bijna alle punten in de gewichtenruimte een rechtlijnig pad te vinden naar een punt met fout nul. De fout zal strikt dalen op dit pad. Dit is echter niet voldoende om te mogen concluderen dat er dan ook een strikt dalend pad is te vinden vanuit de overgebleven verzameling punten (deze heeft maat nul), waar de matrix U singulier is. Dit blijkt uit het voorbeeld dat gegeven is in dit artikel.

Het gevolg is dat bijna alle neurale netwerken lokale minima zullen hebben. De lokale minima, die echter in dit artikel beschreven zijn hebben de eigenschap, dat er vanuit ieder punt buiten het gebied waar U singulier is, er een strikt dalend pad gevonden kan worden naar een absoluut minimum. Er is echter ook een strikt dalend pad naar het lokale minimum toe. Deze lokale minima zijn niet geïsoleerd: als we het gebied bekijken waar zulke lokale minima optreden, dan heeft zo'n gebied randpunten die zadelpunten zijn en van waaruit er weer een directe strikt dalende weg naar het absolute minimum is. Het vraagt om nader onderzoek om te kijken welke leeralgoritmen wel en vooral welke algoritmen niet zullen komen vast te zitten in zo'n lokaal minimum.

5 Literatuur

[1] L.G.C. Hamey; Comments on "Can Backpropagation Error surface Not Have Local Minima?", *IEEE Trans. on Neural Networks*, Vol. 5, No. 5, pp. 844–845, September 1994.

[2] T. Poston, C. Lee, Y. Choie and Y. Kwon; "Local minima and backpropagation", *Proc. of the IEEE-IJCNN91*, vol. II, pp. 173–176, Seattle, 8–12 July 1991.

[3] I.G. Sprinkhuizen-Kuyper and E.J.W. Boers; "The Error Surface of the simplest XOR network has only global minima", *Neural Computation*, Vol. 8, No. 6, 1996 (to appear).

[4] I.G. Sprinkhuizen-Kuyper and E.J.W. Boers; *The Error Surface of the 2-2-1 XOR Network: The finite stationary points*, Technisch Rapport 95-39, RijksUniversiteit Leiden, Vakgroep Informatica.

[5] I.G. Sprinkhuizen-Kuyper and E.J.W. Boers; *The Error Surface of the 2-2-1 XOR Network: Stationary Points with infinite Weights*, Technisch Rapport 96-10, Rijks-Universiteit Leiden, Vakgroep Informatica.

[6] X.H. Yu; "Can backpropagation error surface not have local minima", *IEEE Trans. on Neural Networks*, Vol. 3, No. 6, pp. 1019–1021, November 1992.