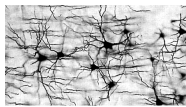


FACULTY OF SCIENCE, MATHEMATICS AND COMPUTING SCIENCE
Nationale Wiskunde Dagen, 3-2-2006
Neurale netwerken

Tom Heskes
IRIS, NIII

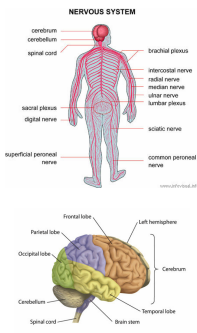


Radboud University Nijmegen

Nationale Wiskunde Dagen, 3-2-2006
Neurale netwerken

Inhoud

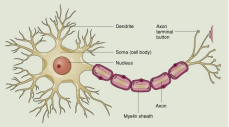

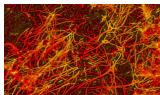
- De bouwstenen van het zenuwstelsel: neuronen en synapsen
- Complex gedrag uit eenvoudige elementen
 - McCulloch-Pitts neuronen
 - Hopfield netwerken
 - Computational neuroscience
- Lerende systemen
 - Perceptrons
 - Meerlaagse perceptrons
 - Toepassing: krantenverkoop



Nationale Wiskunde Dagen, 3-2-2006
Neurale netwerken

Neuronen


- Het brein herbergt zo'n 10.000.000.000 neuronen
- Een neuron is gemiddeld met zo'n 10.000 andere neuronen verbonden

Nationale Wiskunde Dagen, 3-2-2006
Neurale netwerken

Synapsen

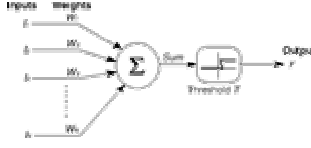
- Communicatie verloopt via synapsen
- Leren = het aanpassen van de sterkte van synapsen / het aanleggen van nieuwe synapsen




Nationale Wiskunde Dagen, 3-2-2006
Neurale netwerken

McCulloch-Pitts neuron

- Spikes zijn 0 of 1
- Sommere gewogen input en drempel



$$s_i(t+1) = \Theta(\sum_j w_{ij}s_j(t) - \mu_i)$$

Nationale Wiskunde Dagen, 3-2-2006
Neurale netwerken

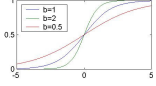
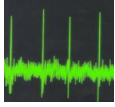
Sigmoids

- Vuurfrequenties i.p.v. individuele spikes:

$$s_i(t+1) = g(\sum_j w_{ij}s_j(t) - d_i)$$
- Of, modelleer ruis:

$$P(s_i(t+1) = 1) = g(\sum_j w_{ij}s_j(t) - d_i)$$
- Typisch sigmoid:

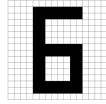
$$g(x) = \frac{1}{1 + \exp(-bx)}$$



Hopfield netwerk

- Associatief geheugen
- Patroonherkenning
- Ruizige M-P neuronen
- Volledig verbonden



- Patronen zijn "bitmaps" met

$$s_i^\mu = 1/0$$

als zwart/wit

- Opslaan door leerregel van Hebb:

$$w_{ij} \propto \sum_{\mu} s_i^{\mu} s_j^{\mu}$$

Hebb's theory

Let us assume that the persistence or repetition of a reverberatory activity (or "trace") tends to induce lasting cellular changes that add to its stability. . . . When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased. — D. O. Hebb (1949)

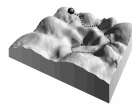


Hopfield netwerk (demo)



Impact

- Conceptueel model van associatief geheugen
- Patroonclassificatie als energie minimalisatie
- Alternatieve methode voor combinatorische optimalisatie
- Neuraal netwerk als statistisch fysisch systeem



Stand van zaken

- Ongeschikt voor toepassingen
- "Computational neuroscience"
- Uitbreidingen:
 - Opslaan van sequenties
 - Rol van oscillaties, synchroniciteit
 - Informatie in spikes ipv frequenties
 - Meer realistische neuron modellen



SINGLE-CELL MODELS: THE COMPARTMENTAL TECHNIQUE
The Hodgkin-Huxley framework

IONIC MOVEMENTS

THE HH EQUATIONS

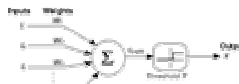
EQUIVALENT CIRCUIT

MODELLED ACTION POTENTIAL



Perceptron

- McCulloch-Pitts neuron als classifier
- Eenvoudige leerregel
- Gegarandeerde convergentie: als er een oplossing bestaat wordt deze gevonden in een eindig aantal stappen



Voorspelling

$$\hat{y} = \Theta(\sum_j w_j x_j - \mu_i)$$

Fout

$$e = \hat{y} - y$$

Update

$$\Delta w_j \propto -e x_j$$

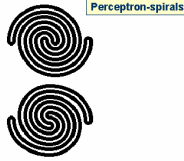


Perceptron (demo)



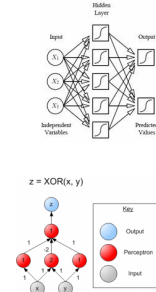
Perceptron

- Minsky & Papert (1969): perceptrons kunnen alleen lineair separeerbare problemen aan
- Veel interessante problemen zijn niet lineair separeerbaar, bijvoorbeeld "connectness"
- "Dark ages" van neurale netwerken



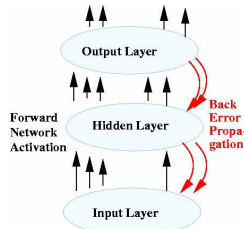
Meerlaags perceptron

- Kan *wel* niet-lineair separeerbare problemen oplossen
- De oplossingen blijken, met relatief eenvoudige aanpassingen, goed te leren
- Interpretatie van verborgen neuronen als gedistribueerde concepten



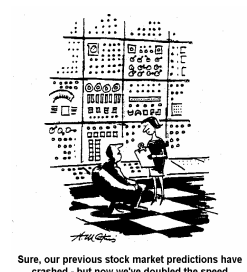
Backpropagation

- Definieer een foutfunctie, bijvoorbeeld het gekwadrateerde verschil tussen de output van het model en de gewenste output
- Bereken de afgeleide van deze fout naar de parameters (gewichten) van het model
- Pas de gewichten aan evenredig met en tegengesteld aan deze afgeleide



Toepassingen

- Veel..., zowel voor classificatie als voor regressie
- Eenvoudig aan de praat te krijgen; veel kant-en-klare software
- Concept trekt ook niet-statistici aan
- Milde niet-lineariteiten komen goed overeen met real-world data



Case

- De Telegraaf:
- 15000 verkooppunten
 - 7 dagen per week

Hoeveel kranten leg je neer op ieder verkooppunt?

te weinig: risico van gemiste verkoop
te veel: onnodige kosten

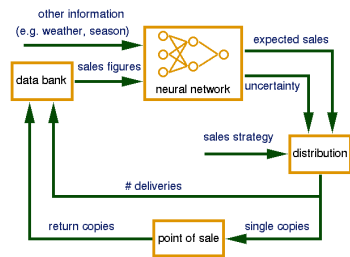


Aanpak

- Regel-gebaseerd
bijv. als uitverkocht, dan verhoog met 1
- Lerend voorspelstelsysteem
 - voorspel de verkoop op verkooppuntniveau
 - schat de nauwkeurigheid van de voorspelling
 - combineer beide met de verkoopstrategie

Idealiter leert dit model continu bij op grond van de meest recente informatie: "neuraal netwerk"

Flow-chart



Verklarende variabelen

Aanname:

Er zijn verklarende variabelen die de verkoop van kranten beïnvloeden en gebruikt kunnen worden om de voorspelling te verbeteren



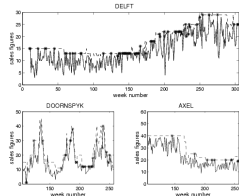
- beter weer, meer verkoop
- slechter weer, meer verkoop

Bijvoorbeeld

- Recente verkoopcijfers
- Verkoopcijfers van vorig jaar
- Uitverkopen
- Weer
- Sportresultaten
- Seizoen en vakanties
-

Overfitten

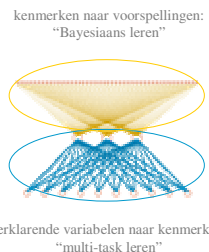
- Typisch per verkooppunt:
 - 100 zeer ruizige data punten (2 jaar)
 - zo'n 20 verklarende variabelen
- Groot risico op "overfitten":
 - prima prestaties op leerset
 - belabberde voorspellingen
- Oplossing: laat verkooppunten "van elkaar leren"!



Van elkaar leren

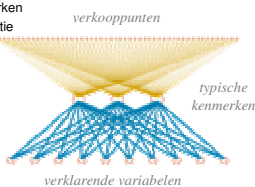
1. Multi-task leren
2. Bayesiaans leren

Doel: verminder het risico op overfitten



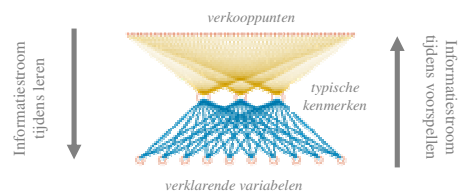
Multi-task leren

- Idee afkomstig uit vakgebied "machine learning"
- Modelleer alle verkooppunten in één "neuraal netwerk":
 - gedeelde vertaling van verklarende variabelen naar typische kenmerken
 - verkooppunt-specifieke combinatie van deze kenmerken
- Reductie van het aantal te schatten parameters per verkooppunt: minder gevaar voor overfitten



Leren en voorspellen

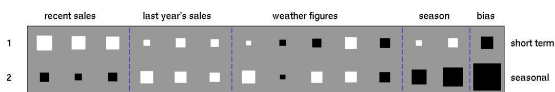
- Streven is het minimaliseren van het kwadratische verschil tussen voorspelde en werkelijke verkoop



- Beide delen kunnen worden gefit met "backpropagation"

Typische kenmerken

Geleerde kenmerken voor een netwerk met 2 hidden units

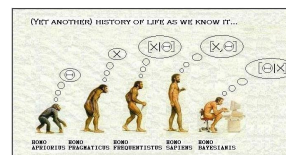


1° hidden unit: gewogen gemiddelde recente verkoop
2° hidden unit: combinatie vorig jaar en seizoen

minder black-box dan vaak wordt beweerd...

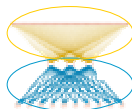
Bayesiaans leren

- Combinatie van a priori information en data: voorkeur voor logische oplossingen
- Automatische regularisatie om overfitten tegen te gaan

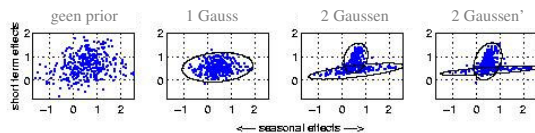


Verschillende priors

zacht gedeeld
verschillend; zelfde prior
"Bayesiaans leren"



hard gedeeld
volledig identiek
"multi-task leren"



Bakker & Heskes, Journal of Machine Learning Research 2003

Iedere dag

- Voor elk verkooppunt:
 - Nieuwe gerealiseerde verkoopcijfers komen binnen.
 - Verkooppunt-specifieke parameters worden bijgeleerd.
 - Nieuwe voorspellingen worden gegenereerd gegeven deze parameters en de verklarende variabelen voor dat moment.
 - Kansmodel geeft daarbij een schatting van de onzekerheid.
- Mogelijke uitkomst: de te verwachten verkoop op maandag voor verkooppunt 1546 is 10.2 ± 2.2 kranten.

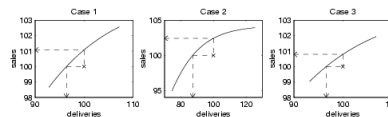
Hoeveel kranten?

- Hoeveel kranten lever je bij voorspelling 10.2 ± 1.2 ?
- En bij 10.2 ± 4.6 ?
- Hangt (mede) af van strategie bedrijf!
- Netjes via (Bayesiaanse) beslissingstheorie



Typische verbeteringen

- In verschillende studies:
 - 1 tot 4% **meer verkoop** bij gelijkblijvend totaal aantal leveringen
 - 3 tot 12% **minder leveringen** voor hetzelfde totaal aantal verkopen




- Door grote aantallen zeer significant en zeer rendabel!



Stand van zaken

- Rekenwerk volledig automatisch
- Oplage manager controleert



- JED 1 sinds 1997 bij 

- Nu: JED 2.0 met veel succes

- Ook implementatie JED 2.0 voor



Take home

- Splitsing in 2 vakgebieden:
 - Computational neuroscience
 - Machine learning, advanced data analysis/statistics
- Computational neuroscience
 - Complexere modellen
 - Gat van neuronmodellen naar imaging en psychofysica
- Machine learning
 - Specifieke methodes voor specifieke problemen
 - Bayesiaans leren vereist goede benaderingen

