1. NEUROVERKKOMENETELMÄT

Ihmisten ja eläinten loistava hahmontunnistuskyky perustuu lukuisiin yksinkertaisiin aivosoluihin ja niiden välisiin kytkentöihin.

Mm. edellisen innoittamana on kehitelty laskennallisia menetelmiä, joita kutsutaan esim. seuraavilla nimillä: '(Artificial) Neural Networks' eli (A)NN, 'Connectionist Modelling' ja 'Parallel Distributed Processing' eli PDP.

Neuroverkkomenetelmien tärkeimmät ominaispiirteet ovat:

- Yksinkertaiset laskentayksiköt, neuronit tai perseptronit
- Verkkomainen struktuuri, synaptiset kytkennät
- Rinnakkainen laskenta
- Black Box -malli syötteen ja vasteen välille

Neuroverkko koostuu (lukuisista) toisiinsa (lokaalisti) kytketyistä neuroneista.

Kytkentöjen vahvuus on säädeltävissä synaptisten painojen avulla.

Yksittäisen neuronin laskentakyky on vähäinen.

Neuronien väliset kytkennät (neuroverkon struktuuri) voivat jakaa monimutkaisen ongelman yksinkertaisiksi osaongelmiksi.

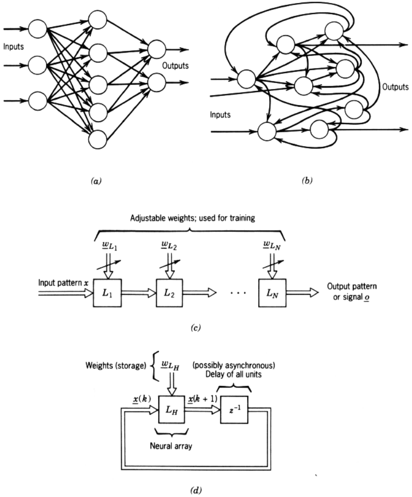
Kokoinaslaskenta-aika nopeutuu, kun osaongelmat ratkaistaan rinnakkaisesti.

Neuroverkot ovat dynaamisia systeemejä, joiden tila (neuronien kytkennät ja ulostulo) muuttuu ajan, sisääntulojen ja alkutilan mukaan.

Hahmontunnistusongelmassa pyritään löytämään neuronien välille sellaiset kytkennät, että syntyy haluttu assosiatiivinen käyttäytyminen: annettu sisääntulo (havainnot) yhdistetään tiettyyn vasteeseen (luokitus).

Klusterointiongelmassa pyritään muodostamaan neuroverkon avulla malli havaintojen rakenteelle (esim. SOM).

Erilaisia neuroverkkojen rakenteita: (a) Feedforward MLP-verkko, (b) rekurentti Hopfield-verkko, (c) Yleinen kaaviokuva (a):sta, (d) yleinen kaaviokuva (b):stä



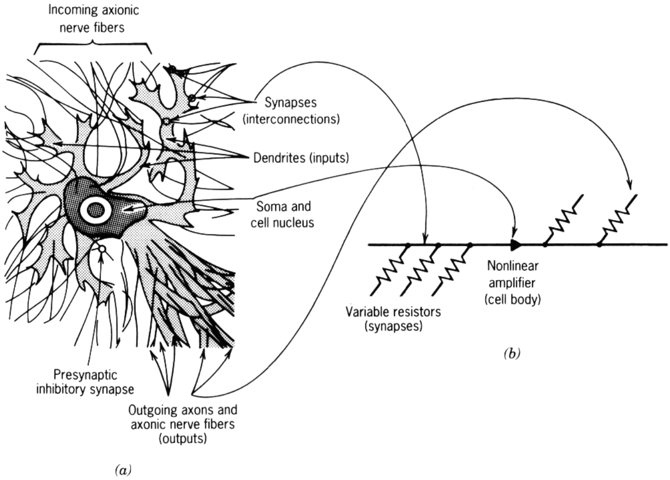
Tyypillisiä piirteitä neuroverkkosovelluksille:

- Korkeaulotteinen piirreavaruus
- Monimutkaiset riippuvuussuhteet piirteiden välillä
- Tyhjä, yksikäsitteinen tai (yleensä) usean tasavertaisen vaihtoehdon muodostama ratkaisujoukko

Hahmontunnistuksessa neuroverkot soveltuvat varsinaisen luokittelun ja vertailun lisäksi esikäsittelyyn ja piirreirrotukseen.

1.1 Perseptroni

Biologinen (a) ja elektroninen (b) neuronimalli:

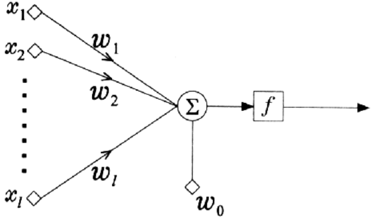


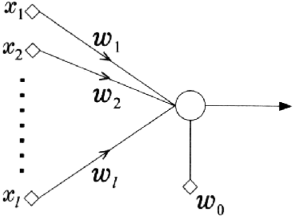
Perseptroni on malli neuroverkon neuronille, joka suorittaa sisääntulolleen $\mathbf{x}$ seuraavan, yksinkertaisen laskutoimituksen:

$$v = w_0 + \sum_{i=1}^{l} w_i x_i = w_0 + \mathbf{w}^T \mathbf{x}, \tag{1}$$

$$y = f(v), \tag{2}$$

missä y on neuronin ulostulo, $\mathbf{w}$ synaptisten kytkentöjen painokertoimet, w_0 kynnysarvo ja $f(\cdot)$ aktivaatiofunktio





Perseptronin aktivaatiofunktio on usein tyypiltään jokin seuraavista:

$$f(x) = \begin{cases} -1, & \text{jos } x < 0 \\ +1, & \text{jos } x > 0 \end{cases}, \tag{3}$$

$$f(x) = \begin{cases} 0, & \text{jos } x < 0 \\ 1, & \text{jos } x > 0 \end{cases}, \tag{4}$$

$$f(x) = \frac{1}{1 + \exp(-ax)}, \tag{5}$$

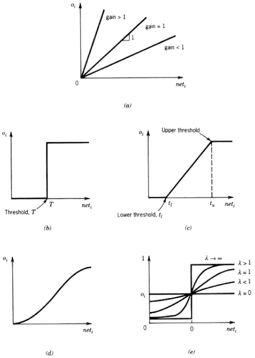
$$f(x) = c\frac{1 - \exp(-ax)}{1 + \exp(-ax)} = c\tanh(\frac{ax}{2}), \tag{6}$$

missä a ja c ovat funktion muotoa sääteleviä parametreja.

Edelliset aktivaatiofunktiot ovat tyypiltään epäjatkuvia askelfunktioita (McCulloch-Pitts perseptroni) tai jatkuvia ja derivoituvia litistysfunktioita (squashing function).

Molemmilla tyypeillä on puolensa: askelfunktioilla neuroverkon suorittaman kuvauksen analysointi, ja jatkuvilla funktiolla neuroverkon opettaminen on helpompaa.

Erilaisia aktivaatiofunktiota: (a) lineaarinen, (b) kynnys/askelfunktio, (c) lineaarinen kynnysfunktio, (d) sigmoidi, (f) sigmoideja erilaisilla parametrin arvoilla



Kun käytetään em perseptroneja diskriminanttifunktiona, voidaan piirreavaruus jakaa lineaarisesti kahteen osaan.

Yksittäinen perseptroni suoriutuu siis melko yksinkertaisesta luokitteluongelmasta.

(Perseptronin opettamisesta puhuttiin lineaaristen luokittimien yhteydessä.)

(Aktivaatiofunktiot voivat olla myös RBF-tyyppisiä.)

1.2 MLP-verkko

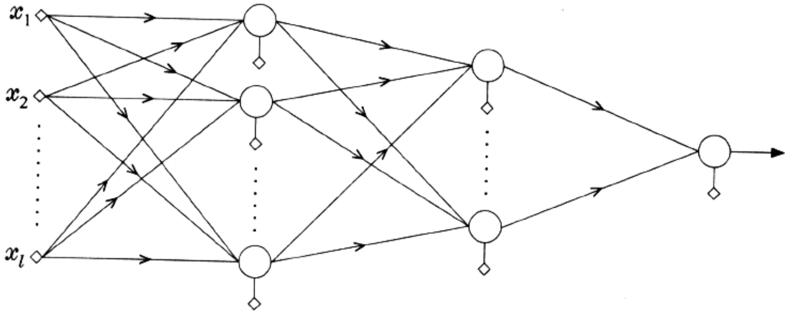
Kytkemällä perseptroneja toisiinsa, saadaan aikaiseksi monimutkaisemmista ongelmista suoriutuva neuroverkko.

Eräs tyypillinen ratkaisu on kytkeä perseptronit toisiinsa siten, että ei synny silmukoita (feedforward network).

Tämän verkkoarkkitehtuurin yksi yleinen erikoistapaus on monikerrosperseptroni (multi-layer perceptron, MLP).

MLP-verkossa perseptronit on järjestetty useiksi kerroksiksi, joiden sisällä ei ole kytkentöjä. Kerrokset on järjestetty ja jokainen kerros kytketty täydellisesti edeltäjäänsä ja seuraajaansa.

Kuva 3-kerrosperseptronista, jolla on yksi ulostuloperseptroni:



MLP:n 1. kerros on sisääntulokerros, jossa on jokaiselle piirteelle oma perseptroni. Sisääntulokerroksen perseptronit eivät suorita lainkaan laskentaa.

Viimeinen kerros on ulostulokerros, jonka perseptronien ulostulo on koko neuroverkon ulostulo eli vaste. Ulostuloperseptronien aktivaatiofunktio on usein lineaarinen.

Ulostuloperseptronien lkm riippuu ongelmasta.

Väliin jääviä kerroksia kutsutaan piilokerroksiksi. Piiloperseptronien aktivaatiofunktion tyyppi on usein kaikille sama (jos tyyppi on RBF, verkkoa kutsutaan RBF-verkoksi).

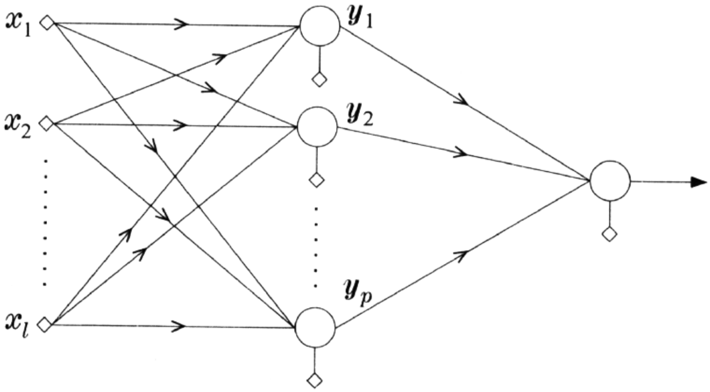
Sopiva piiloperseptronien lkm riippuu ongelmasta.

Tarkastellaan seuraavaksi 2- ja 3-kerroksisia MLP-verkkoja, joiden perseptronit on McCulloch-Pitts-tyyppisiä.

2-kerrosperseptroni

('Two-layer perceptron')

Yksinkertaisimmillaan MLP-verkossa on vain yksi piilokerros ja yksi ulostuloperseptroni (vastaa kahden luokan ongelmaa):



Seuraava tarkastelu on helppo yleistää useammalle ulostuloperseptronille (useamman kuin kahden luokan ongelma).

Voidaan ajatella, että 2-kerrosperseptroni ratkaisee luokitteluongelman kahdessa perättäisessä vaiheessa.

Piilokerros määrittää piirreavaruuteen hypertasoja ja laskee sisääntulon aseman niihin nähden.

Ulostulokerros yhdistelee piilokerroksen perseptronien ulostuloja ja päättelee mihin osaan piirreavaruutta ja mihin luokkaan havainto kuuluu.

Voidaan myös ajatella, että piilokerros suorittaa epälineaarisen kuvauksen uuteen piirreavaruuteen ja ulostulokerros suorittaa varsinaisen lineaarisen luokittelun (vrt. viime luonnolla esitelty yleistetty lineaarinen luokitin).

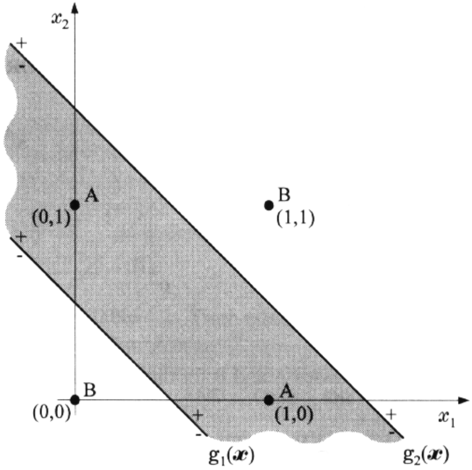
Esimerkki: XOR ja 2-kerrosperseptroni

Ratkaistaan 2-ulotteinen XOR-ongelma käyttäen 2-kerrosperseptronia.

Perseptronien aktivaatiofunktiot saavat joko arvon 0 (epätosi) tai 1 (tosi).

Käytetään piilokerroksessa kahta perseptronia.

Valitaan piiloperseptronien painokertoimet siten, että tosi-luokan havainnot jäävät perseptronien määrittämien suorien väliin ja epätosi-luokan havainnot taas niiden ulkopuolelle:



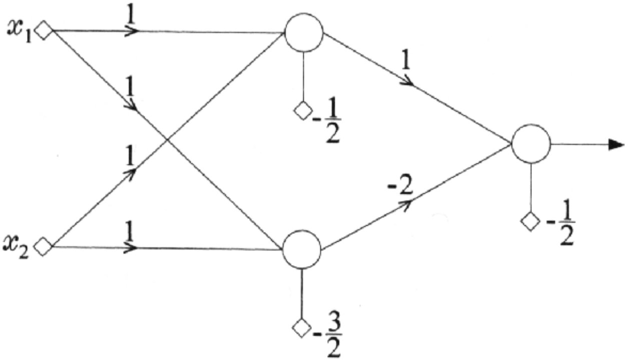
Valitaan ainoan ulostuloperseptronin painokertoimet siten, että verkon ulostulo on 0 tai 1 seuraavan taulukon mukaisesti:

x_1	x_2	y_1^h	y_2^h	y^o
0	0	0	0	0
0	1	1	0	1
1	0	1	0	1
1	1	1	1	0

missä x_1 ja x_2 ovat verkon sisääntulot, y_1^h ja y_2^h piiloperseptronien ulostulot ja y^o verkon ulostulo.

Toinen piiloperseptroni suorittaa AND- ja toinen OR-operaation (vrt. aikaisempi esimerkki).

Verkon painokertoimet voidaan valita esimerkiksi seuraavasti:



2-kerrosperseptronin luokittelukyky

Tarkastellaan seuraavaksi millaisista kahden luokan ongelmista 2-kerrosperseptroni selviytyy (tulokset yleistyvät helposti useammalle kuin kahdelle luokalle).

Ol., että aktivaatiofunktiot ovat kaksiarvoisia (0 tai 1).

Ol., että verkossa on l inputperseptronia, p piiloperseptronia ja yksi ulostuloperseptroni.

Piilokerros kuvaa piirrevektorit p-ulotteisen hyperkuution H_p kulmiin:

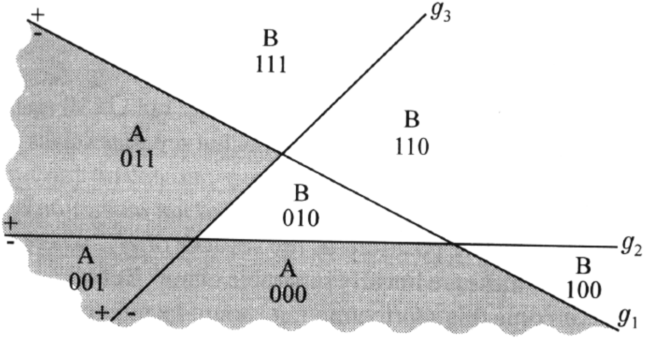
$$H_p = \{[y_1, \ldots, y_p]^T, y_i \in [0, 1], 1 \leq i \leq p\} \tag{7}$$

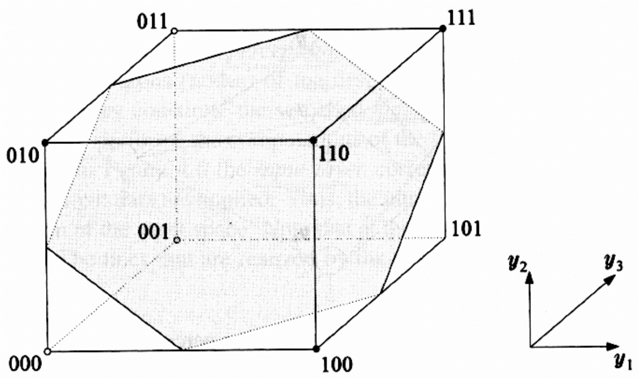
Hyperkuution kulmia ovat ne pisteet, joissa y_i:iden arvot ovat joko nollia tai ykkösiä.

Jokainen piiloperseptroni vastaa hypertasoa. Nämä hypertasot jakavat piirreavaruuden osiin, monitahokkaiksi, joista jokainen vastaa yhtä hyperkuution H_p kulmaa.

Ulostuloperseptroni jakaa hyperkuution H_p lineaarisesti kahteen osaan. Eri hyperkuution osiin jäävät kulmat vastaavat eri luokkia.

2-kerrosperseptronin muodostamat päätösalueet ovat siis monitahokkaiden unioneita.




011
111
010
110
001
101
000
100
y_2
y_3
y_1

3-kerrosperseptroni

Tarkastellaan seuraavaksi 3-kerrosperseptronia, jossa on 2 piilokerrosta ja joka pystyy suorittamaan monimutkaisemman kuvauksen kuin 2-kerrosperseptroni.

Tarkastellaan kahden luokan ongelmaa.

Ol. että perseptronien ulostulot ovat kaksiarvoisia.

Ol., että luokkaan ω_1 kuuluvat kaikki havainnot, jotka sattuvat tiettyihin (J kpl) 1.:n piilokerroksen p:n perseptronin määrittämiin monitahokkaisiin. Luokkaa ω_2 vastaavat kaikki loput monitahokkaat.

2. piilokerros pystyy muodostamaan kaikki mahdolliset J:n monitahokkaan unionit, jos siinä on J piiloperseptronia

Todistuksen luonnehdinta:

- 1. piilokerroksen perseptronit määrittävät p hypertasoa piirreavaruudessa.

- 1. piilokerros kuvaa piirrevektorit p-ulotteisen hyperkuution kulmille. Jokainen kulma vastaa yhtä hypertasojen rajaamaa monitahokasta piirreavaruudessa.

- Valitaan 2.:n piilokerroksen perseptronien painot siten, että yksi perseptroni erottaa yhden hyperkuution kulman kaikista muista.

- Valitaan eroteltavat kulmat siten, että ne vastaavat luokkaa ω_1.

- Valitaan ulostuloperseptronin painot siten, että perseptroni toteuttaa OR-operaation.

- Nyt verkon ulostulo on tosi (1), jos piirrevektori sattuu johonkin luokkaa ω_1 vastaavaan monitahokkaaseen ja muulloin epätosi (0).

2. piilokerrokseen ei välttämättä tarvita J:tä piiloperseptronia - minimimäärä riippuu ongelmasta.

Yhteenveto: 1. piilokerros määrittää hypertasot piirreavaruuteen, 2. piilokerros määrittää hypertasojen rajaamien monitahokkaiden unioneita, ja ulostulokerros määrittää eri luokkia vastaavat päätösalueet.

(Jos perseptronien aktivaatiofunktiot eivät ole askelfunktioita, MLP-verkon toiminnan tulkinta ei ole näin suoraviivaista.)

1.3 Universaali approksimaattori

Edellisten tarkastelujen perusteella MLP-verkko on selvästi epälineaarinen luokittelumenetelmä.

MLP-verkko, jonka piiloperseptronit ovat epälineaarisia ja ulostuloperseptronit lineaarisia ja jolla on yksi ulostuloperseptroni, on yleistetty lineaarinen luokitin (ks. edellinen luento).

Tarkastellaan seuraavaksi tällaisen MLP-verkon ja muiden yleistettyjen lineaaristen luokittimien approksimointikykyä.

Voidaan ajatella, että yleistetty lineaarinen luokitin approksimoi diskriminanttifunktioita (tai luokkien *a posteriori* tnjakaumia) epälineaaristen kantafunktioiden painotetulla summalla.

Lähdetään aluksi liikkeelle **Weierstrassin teoreemasta**:

- Olkoot $g(\mathbf{x})$ jatkuva funktio, joka on määritelty piirreavaruuden suljetussa osajoukossa S.

- Olkoot $\epsilon > 0$ jokin mielivaltaisen pieni vakio.

- On olemassa $r(\epsilon)$ ja astetta r oleva polynomi $\Phi(\mathbf{x})$ siten, että

$$|g(\mathbf{x}) - \Phi(\mathbf{x})| < \epsilon, \; \forall \mathbf{x} \in S \tag{8}$$

- Ts. jatkuvaa funktiota voidaan approksimoida mielivaltaisen tarkasti polynomilla (suljetussa) kompaktissa alueessa, kunhan polynomi on riittävän korkeaa astetta.

Suurin ongelma polynomilla approksimoinnissa on se, että usein riittävän tarkkuuden saavuttamiseksi tarvitaan korkea asteluku r.

Korkeaulotteisten polynomien ongelmana on laskennan vaativuuden lisäksi huono yleistämiskyky ja mittausvirheiden kertautuminen.

Ongelmaa voidaan kiertää approksimoimalla funktiota paloittain polynomeilla.

Hidas approksimointivirheen pienentyminen on ongelma myös muille yleistetyille lineaarisille luokittimille, joiden kantafunktiot ovat kiinteitä.

Mikäli kantafunktiot eivät ole kiinteitä vaan määritelty opetusjoukon perusteella säädettävien parametrien avulla (vrt. MLP, RBF), approksimaatiovirhe pienenee paljon nopeammin.

Tarkastellaan 2-kerrosperseptronia, jonka ainoan ulostuloperseptronin vaste voidaan kirjoittaa seuraavasti:

$$\Phi(\mathbf{x}) = \sum_{j=1}^{k} w_j^o f(\mathbf{x}^T \mathbf{w}_j^h + w_{j0}^h) + w_0^o \tag{9}$$

missä k on piiloperseptronien lkm ja painokertoimen yläindeksit o ja h viittaavat piilo- ja ulostulokerroksiin

Kun lisäksi oletetaan, että $f(\cdot)$ on litistysfunktio, on vasteella $\Phi(\mathbf{x})$ seuraavat **universaalit approksimointiominaisuudet**:

- Olkoot $g(\mathbf{x})$, jatkuva funktio, joka on määritelty piirreavaruuden kompaktissa osajoukossa S
- Olkoot $\epsilon > 0$ jokin mielivaltaisen pieni vakio
- On olemassa $k(\epsilon)$ ja k:n piiloperseptronin 2-kerrosperseptroni siten, että
$$|g(\mathbf{x}) - \Phi(\mathbf{x})| < \epsilon, \; \forall \mathbf{x} \in S \tag{10}$$

- Approksimointivirhe pienenee säännön $O(\frac{1}{k})$ mukaisesti (Ei riipu piirreavaruuden ulottuvuudesta!)

- Tämä tulos pätee myös RBF-verkolle

- Ts. jatkuvaa funktiota voidaan approksimoida mielivaltaisen tarkasti MLP- tai RBF-verkolla kompaktissa alueessa, kunhan perseptroneja (kantafunktiota) on riittävän monta.

MLP-approksimoinnin haittapuolena on se, että sopivien painokertoimien valinta on epälineaarinen optimointiongelma (lokaalit minimit!).

Mitä hyötyä on käyttää MLP-verkkoa, jossa on useampi kuin yksi piilokerros? Silloin approksimointi on tehokkaampaa ja tietyn tarkkuuden saavuttamiseksi tarvitaan yleensä vähemmän perseptroneja.

1.4 Neuroverkon opettaminen

MLP-verkot, kuten muutkin neuroverkot tai yksittäiset perseptronit, opetetaan opetusjoukon avulla.

Opettamisella tarkoitetaan yleensä sopivien synaptisten painojen ja kynnysarvojen (verkon vapaiden parametrien) määräämistä.

Laajemmin ajateltuna oppimisena voidaan pitää myös verkon struktuurin valintaa.

Oppiminen voi olla joko ohjattua (esim. backpropagation), ohjaamatonta (esim. Hebbian, Self-Organizing Map eli SOM) tai vahvistettua (reinforcement learning).

Ohjaamattomassa oppimisessa yritetään löytää havainnoille mielekäs vaste:

- (anti)Hebbian-oppiminen vahvistaa niiden neuronien välisiä kytkentöjä, jotka aktivoituvat yhtä(eri)aikaisesti. Hebbian-tyyppisiä synapseja kutsutaankin korrelaatiosynapseiksi

- On olemassa fysiologisia todisteita sille, että hippokampuksessa (aivoturso), jolla on tärkeä rooli oppimisessa ja muistamisessa, tapahtuu Hebbian-tyyppistä oppimista.

- SOM yrittää mallintaa havaintojen rakennetta neuronihilan avulla siten, että lähellä toisiaan olevat neuronit aktivoituvat samankaltaisista havainnoista (tästä lisää klusteroinnin yhteydessä).

Ohjatussa oppimisessa haluttu verkon vaste (luokka) tunnetaan. Verkon struktuuri ja synaptiset painot pyritään valitsemaan siten, että verkon todellinen vaste on mahdollisimman lähellä sen haluttua vastetta ja että verkon struktuuri on mahdollisimman yksinkertainen.

Ohjatun oppimisen menetelmät voidaan jakaa kahteen luokkaan: 1) opetusjoukon täydelliseen luokitteluun perustuvat menetelmät ja 2) verkon halutun ja todellisen vasteen eroa kuvastavan kustannusfunktion minimointiin perustuvat menetelmät.

Opetusjoukon täydelliseen luokitteluun perustuvat menetelmät

Näissä menetelmissä lähdetään liikkeelle yksinkertaisesta neuroverkosta, joka ei yleensä pysty ratkaisemaan annettua tehtävää.

Verkkoon lisätään perseptroneja, kunnes se hallitsee opetusjoukon täydellisesti.

Lähtökohtana on ajatus ongelman jakamisesta helpompiin osaongelmiin, jotka pystytään ratkaisemaan yhdellä perseptronilla tai hyvin yksinkertaisella verkolla (constructive techniques).

Verkon parametrit voidaan määrätä lineaaristen luokittimien yhteydessä esitetyillä menetelmillä (esim. perseptroni- tai LMS-algoritmillä).

Verkon asteittaiseen kasvattamiseen on kehitelty useita menetelmiä, jotka perustuvat uusien piilokerrosten tai piiloperseptronien lisäämiseen. Osa menetelmistä tuottaa MLP-verkkoja, mutta osa sallii kytkennät muidenkin kuin perättäisten kerrosten välillä tai kytkennät kerrosten sisällä.

Esimerkki: Tiling-algoritmi

Tiling-algoritmi on yksi konstruktiivinen menetelmä.

Se tuottaa MLP-verkkoja, joissa on useita kerroksia, yleensä vähintään kolme.

Tarkastellaan kahden luokan ongelmaa:

- Algoritmi lähtee liikkeelle yhdestä master-perseptronista, joka on ensimmäisessä kerroksessa ja joka opetetaan pocket-algoritmillä
- Jaetaan opetusjoukko X kahteen osajoukkoon: päätöspinnan positiiviselle (X^+) ja negatiiviselle puolelle (X^-) sijoittuvat havainnot
- Mikäli jompikumpi tai molemmat osajoukot sisältävät näytteitä molemmista luokista, lisätään ensimmäiseen kerrokseen apu-perseptronit: $n(X^+)$ ja/tai $n(X^-)$, jotka opetetaan osajoukoilla X^+ ja/tai X^-
- Jaetaan osajoukot X^+ ja X^- tarvittaessa uusiksi osajoukoiksi X^{++}, X^{+-}, X^{-+}, ja X^{--} ja lisätään niitä vastaavat apu-perseptronit

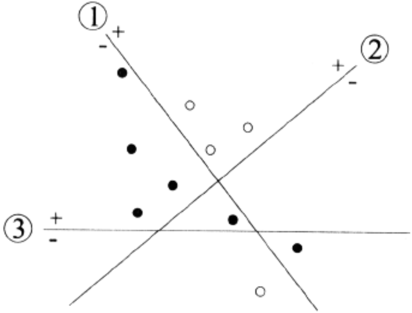
- Lisätään apu-perseptroneja kunnes ei pystytä muodostamaan uusia osa-joukkoja

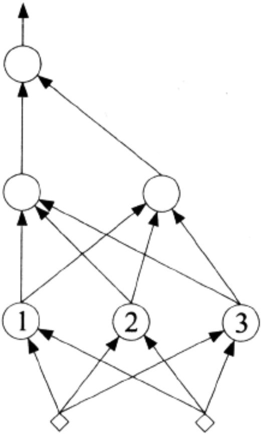
- Lisätään uusi kerros ja siihen master-perseptroni, joka saa syötteensä kaikilta edellisen kerroksen perseptroneilta ja jota opetetaan koko opetusjoukolla. Kasvatetaan uutta kerrosta apu-perseptroneilla kuten ensimmäistä kerrosta

- Lisätään uusia kerroksia, kunnes pystytään luokittelemaan kaikki opetusnäytteet oikein

- Voidaan osoittaa, että verkkoon lisätään vain äärellinen määrä kerroksia (jokainen uusi kerros pystyy luokittelemaan oikein kaikki ne näytteet kuin edellisen kerroksen master-perseptronikin ja vähintään yhden näytteen enemmän)

Luokitteluongelman ratkaisu Tiling-algoritmillä:





Esimerkki: lähimmän naapurin menetelmään perustuva verkko

Toinen esimerkki konstruktiivisesta oppimismenetelmästä on MLP-verkon rakentaminen lähimmän naapurin menetelmään (kNN) perustuen:

- MLP-verkon 1. piilokerros koostuu perseptroneista, jotka määrittävät hypertason jokaisen opetusnäyteparin välille
- 2. piilokerros muodostaa hypertasoista monitahokkaita AND-perseptronien avulla
- Ulostulokerros muodostaa päätösalueet OR-perseptronien avulla

Menetelmän haittapuolena on se, että verkkoon tarvitaan runsaasti neuroneja.