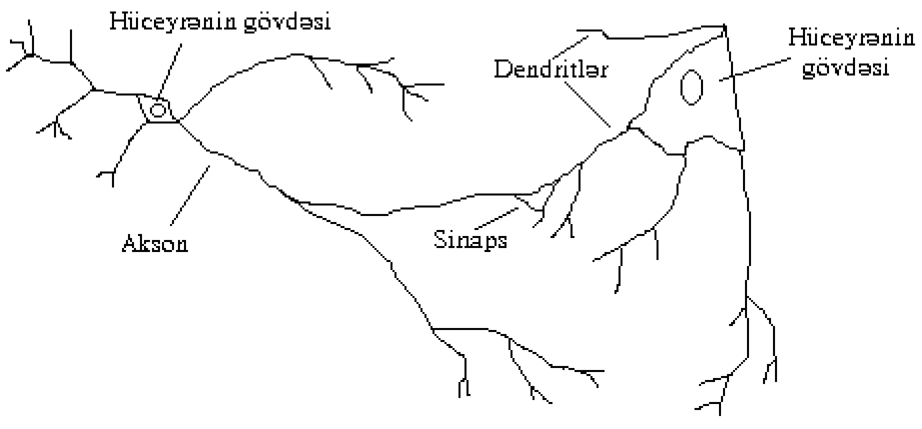
6 NEYROKOMPÜTİNQ

6.1 Neyron və Neyron Şəbəkələri

Bioloji neyron. Əsəb hüceyrələrinin müxtəlif növləri ümumi neyron termini altında birləşdirilir. Neyron canlı orqanizmlərdə əsas məqsədi orqanizmin operativ idarə edilməsi olan elektrik aktivliyə malik xüsusi şəkilli hüceyrə kimi təsəvvür edilir. Bioloji neyronun sxematik təsviri şəkil 3.1-də göstərilmişdir. Neyron örtüklə – membranla əhatə olunmuş nüvə və sitoplazmadan təşkil edilən hüceyrə gövdəsindən ibarətdir. Neyrona həm də girişlər (dendritlər) və çıxışlar (akson və onun ucları) ağacı daxildir. Aksonun son şaxələri (budaqları) başqa hüceyrələrə sinaptik əlaqələrlə (sinapslarla) birləşir. Aksondan həm də başqa hüceyrələri birləşdirən çıxıntılar – kollaterallar çıxır.



Şəkil 6.1 Bioloji neyron

Dendrit ağacının giriş siqnalları (postsinaptik potensiallar) aksonun başlanğıc seqmentinə doğru yolda cəmlənir və nəticədə çıxış impulsu (və ya impulslarlı) generasiya olunur. Beləliklə onun intensivliyi giriş siqnallarının əmsallaşdırılmış cəmindən asılı funksiya olur. Çıxış siqnalı aksonun budaqlarından keçir və digər neyronların dendrit ağaclarını aksonla birləşdirən sinapslara çatır. Siqnal sinapslar vasitəsilə qonşu neyronlar üçün yeni giriş siqnalına çevrilir. Sinapsların növündən asılı olaraq bu siqnal müsbət və mənfi (həyəcanlandırıcı və ləngidici) ola bilər. Sinapsın generasiya etdiyi giriş siqnalının qiyməti sinapsa gələn siqnal eyni olduqda belə müxtəlif ola bilər. Bu müxtəliflik sinapsın effektivliyi ilə əlaqədardır ki, bu onun çəki əmsalı ilə müəyyən edilir. Sinapsın fəaliyyəti prosesində onun çəki əmsalı dəyişilə bilər.

Neyronlar reseptor, aralıq və effektor tipli üç qrupa bölünürlər. Reseptor neyronlar beyinə sensor informasiyanın daxil olmasını təmin edir. Onlar hissiyyat üzvlərinə gələn siqnalları (göz torlarında optik, qulaq hələzununda akustik siqnallar və s.) öz aksonlarının elektrik pulsasiyasına çevirirlər. Effektor neyronlar onlara daxil olan siqnalları icra orqanlarına ötürürlər. Onların aksonlarının axırında icra orqanları ilə xüsusi sinaptik birləşmələr olur, məsələn, əzələlərlə, burada neyronların həyəcanlanması (*oyanması*) əzələlərin yığılmasına çevrilir. Aralıq neyronlar reseptorlardan alınan informasiyanın emal edilməsini həyata keçirir və effektorlar üçün idarəedici siqnallar yaradırlar.

Bir neyronun sinapslarının sayı yüzdən minə qədər dəyişir. Neyronun membranının vəziyyəti (onun elektrik potensialı) sinapslara daxil olan siqnalların sürəkliliyindən (uzunluğundan) və qiymətindən asılıdır. Membranın potensialı müəyyən hədd qiymətinə (təxminən 40 mV) çatanda əsəb impulsu – aksonun əsəb teli boyunca yayılan fəallaşma dalğası əmələ gəlir. Əsəb teli üzrə ötürülən qıcıqlanma elektrokimyəvi proses olub yayılma sürəti telin diametrindən asılıdır və $1\ \text{m/san} \div 150\ \text{m/san}$ arasında dəyişir. İmpulsu ötürdükdən sonra əsəb teli tam sakitlik (refraktor dövrü) vəziyyətində olur, qıcıqlanma təsirinin hansı dərəcədə olmasından asılı olmayaraq əsəb siqnallarını ötürmür.

Neyronun riyazi modeli. Neyronun, çoxlu sayda qarşılıqlı əlaqədə olan neyronlardan ibarət neyron şəbəkələrinin modelləşdirilməsində adətən istifadə olunan riyazi modeli şəkil 6.2 göstərilmişdir [1,2,3]. Neyronun $x_1, x_2, \ldots, x_n$ giriş siqnalları toplusu (və ya X giriş vektoru) digər neyronların çıxış siqnallarından təşkil olunur. Bu giriş vektoru bioloji neyronların sinapslarına daxil olan siqnallara uyğundur. Hər bir giriş siqnalı uyğun $w_1, w_2, \ldots, w_n$ əlaqə çəkilərinə (sinapsın effektivlik analoqu) vurulur. Əlaqə çəkiləri skalyar kəmiyyətdir, həyəcanlandırıcı əlaqələr üçün müsbət və ləngidici əlaqələr üçün mənfidir. Əlaqə çəkilərinə vurulmuş giriş siqnalları hüceyrənin gövdəsinə uyğun olan cəmləmə blokuna daxil olur. Burada onların cəbri cəmlənməsi həyata keçirilir və neyronun oyanma səviyyəsi müəyyən edilir

$$I = \sum_{i=1}^{n} w_i x_i \tag{6.1}$$

Neyronun y çıxış siqnalı I oyanma səviyyəsinin qeyri-xətti F funksiyası vasitəsilə çevrilməsi kimi təyin olunur

$$y = F(I - \theta), \tag{6.2}$$

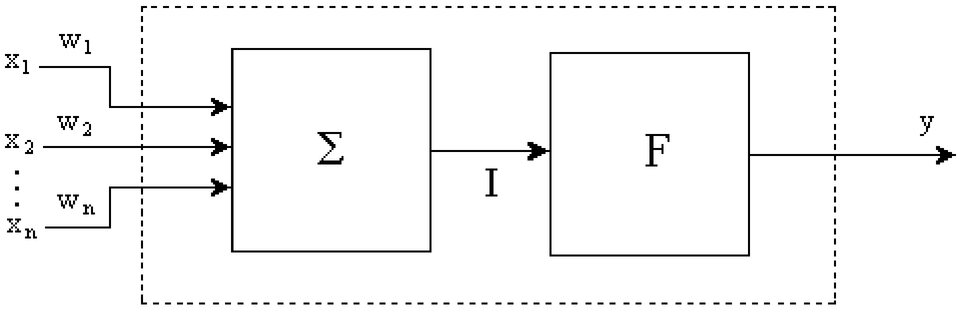
burada θ – neyronun oyanma hüdududur. Adətən F funksiyası olaraq aşağıdakı qeyri-xətti funksiyalar götürülür:

a) vahid sıçrayış funksiyası

$$y = \begin{cases} 1, & I > \theta \quad \text{олдугда} \\ 0, & I \le \theta \quad \text{олдугда} \end{cases}$$

b) siqmoid

$$y = \frac{1}{1 - e^{-(I-\theta)}}$$



Şəkil 6.2 Neyronun riyazi modeli

Neyronları bir-biri və xarici mühitlə müəyyən qaydada birləşdirərək onların yuxarıda göstərilmiş modelləri əsasında neyron şəbəkəsini qurmaq olar. Giriş siqnal vektoru şəbəkəyə giriş neyronlarının fəallaşdırılması yolu ilə verilir. Şəbəkənin neyronlarının $y_1, y_2, \ldots, y_n$ çıxış siqnalları çoxluğu çıxış fəallığı vektoru adlanır. Şəbəkənin neyronlarının əlaqə çəkiləri elementləri i -ci və j -cu neyronların w_{ij} əlaqə çəkiləri olan W matrisi kimi göstərilir. Şəbəkənin fəaliyyəti prosesində giriş vektorunun çıxış vektoruna çevrilməsi, yəni informasiyanın işlənməsi həyata keçirilir. Qeyd etmək lazımdır ki, neyron şəbəkə vasitəsi ilə yerinə yetirilən informasiya çevrilməsi yalnız neyron modellərinin xarakteristikaları ilə deyil, həm də onun arxitekturası ilə şərtlənir.

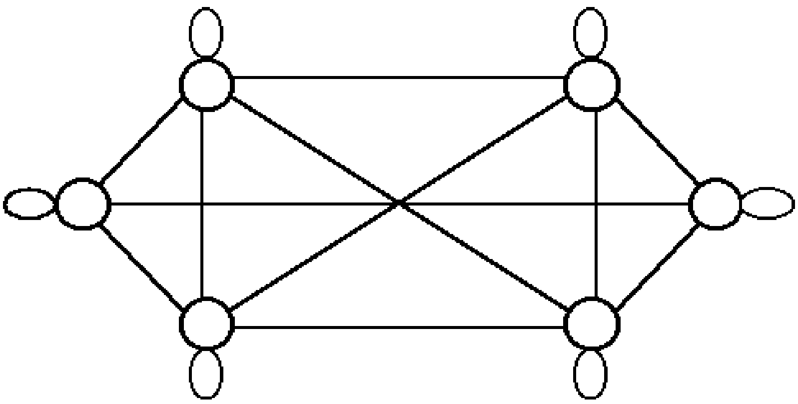
Neyron şəbəkələrinin arxitekturası və tiplərinə baxaq.

Tam əlaqəli və iyerarxik şəbəkələr. Neyron şəbəkə müəyyən bir struktur ilə əlaqələndirilmiş ayrı-ayrı neyronlar toplusudur. Şəbəkənini hesablama gücü, yəni onun yerinə yetirə biləcəyi məsələlər məhz bu əlaqələrlə müəyyənləşir. Əlaqələr bir qisim neyronların çıxışlarını digərlərinin girişləri ilə birləşdirir, onların "gücü" isə çəki əmsalları ilə (və ya sadəcə olaraq çəkilərlə) verilir. Beləliklə, bir neyronun davranışının digər neyronun davranışına təsir gücü əlaqənin uyğun çəki əmsalı ilə müəyyən olunur. Buna görə də neyron sistemlərini çox vaxt konneksionist (*connection* – birləşmə, əlaqə sözündən) sistemlər də adlandırırlar.

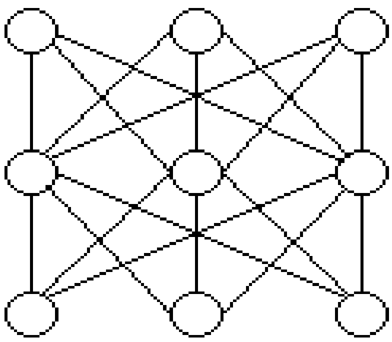
Şəbəkədə əlaqələrin düzümü onun arxitekturasını təşkil edir [1,2,36]. Neyron şəbəkələrin arxitekturasını iki növə ayırmaq olar: tam əlaqəli və iyerarxik şəbəkələr.

Qraflar nəzəriyyəsindən məlumdur ki, tam əlaqəli arxitekturada şəbəkənin bütün elementləri bir-biri ilə birləşmiş olur. Neyron şəbəkələrdə bu onu göstərir ki, hər bir neyronun çıxışı bütün digər neyronların girişləri ilə birləşmiş və onun girişləri qalan elementlərin çıxışları ilə bağlanmışdır. Bundan başqa, hər bir neyronun çıxışı onun girişinə birləşdirilmişdir ("özü-özüylə əlaqə"). N neyrondan ibarət tam əlaqəli şəbəkələrdə hər bir düyündən N əlaqə çıxdığından əlaqələrin sayı $N \cdot N$-ə bərabərdir. Şəkil 6.3-də 6 neyrondan ibrət olan tam əlaqəli şəbəkə göstərilmişdir.

İyerarxik arxitekturada ayrı-ayrı laylarda və ya səviyyələrdə yerləşmiş neyron qruplarını ayırmaq olar. Layın hər bir neyronu əvvəlki və sonrakı layın hər bir neyronu ilə əlaqələndirilir.



Şəkil 6.3 Tam əlaqəli neyron şəbəkəsi



Şəkil 6.4 İyerarxik neyron şəbəkəsi

Giriş və çıxış laylarını xüsusi olaraq ayırmaq lazımdır. Giriş layının neyronları siqnalları xarici mühitdən alırlar və sonrakı layın neyronları üzrə paylayırlar. Çıxış layının neyronlarının çıxışları xarici mühitə daxil olur. Girişlər və çıxışlar arasında yerləşmiş laylar aralıq və yaxud gizli laylar adlanır (onların xarici mühitlə bilavasitə əlaqəsi olmur). Şəkil 6.4-də üç laylı iyerarxik neyron şəbəkəsi təsvir edilmişdir.

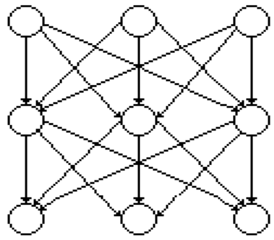
Siqnalların ötürürülməsi istiqamətinə görə şəbəkələri əks əlaqəsiz və ya qeyri-rekurent (feed-forward) və əks əlaqəli və ya rekurent (*feed-back*) şəbəkələrə ayırmaq olar.

Əks əlaqəsiz şəbəkələrdə bir layın neyronları siqnalları yalnız xarici mühitdən və yaxud da əvvəlki layın neyronlarından qəbul edirlər və siqnalları ya xarici mühitə, ya da sonrakı layın neyronlarının girişinə ötürürlər.

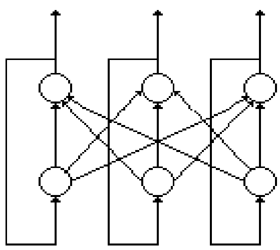
Rekurent şəbəkələrdə müəyyən laydakı neyronlar bundan əlavə, siqnalları özləri-özlərindən və həmin layda yerləşən digər neyronlardan qəbul edə bilərlər. Beləliklə, qeyri-rekurent şəbəkələrdən fərqli olaraq rekurent şəbəkənin neyronlarının çıxış siqnallarının qiyməti yalnız girişlərindəki siqnalların cari qiymətləri və uyğun əlaqələrin çəkiləri ilə deyil, həm də bəzi neyronların əvvəlki zaman anındakı çıxış qiymətləri ilə təyin olunurlar. Bu onu göstərir ki, belə şəbəkə çıxışların vəziyyəti haqqında informasiyanı müəyyən müddətdə yadda saxlamağa imkan verən yaddaş elementlərinə malikdir.

Rekurent şəbəkələrin öz layının neyronları ilə əlaqələri ləngidici olan halda (yəni mənfi çəkili əlaqələr olan hal) onu lateral ləngidici şəbəkə adlandırırlar. Şəkil 6.5, 6.6 və 6.7-də uyğun olaraq əks əlaqəsiz, rekurent və lateral ləngidici şəbəkələr təsvir edilmişdir.

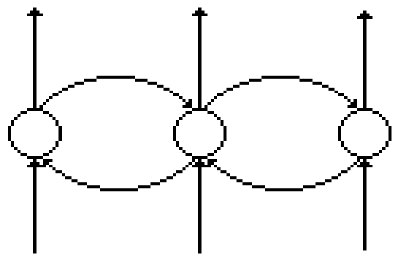
Xarici mühitlə ən sıx qarşılıqlı əlaqə supervizor üsüllarından istifadə edəndə olur. Bu zaman giriş vektorlar çoxluğu və ona uyğun çıxış vektorları çoxluğu əvvəlcədən müəyyənləşdirilir. Hər bir giriş vektorunun i - ci komponenti şəbəkənin i - ci neyronuna verilən siqnala uğun gəlir. Analoji olaraq, çıxış vektorunun j - cu komponenti j - cu çıxış neyronunda alınan siqnala uyğun gəlir. X giriş vektoru və ona uyğun Y çıxış vektoru öyrətmə cütləri adlanan cüt təşkil edir. Bütün öyrətmə cütləri toplusu öyrətmə çoxluğunu yaradır. Öyrətmə prosesi zamanı neyron şəbəkənin çıxış vektorunun cari qiymətləri ilə öyrətmə çoxluğundan seçilmiş verilmiş qiymətlər arasındakı meyletmələr hesablanır. Bu qiymətləndirməyə uyğun olaraq şəbəkənin parametrlərində təshih aparılır. Neyron şəbəkələrin öyrətmə alqoritmləri arasında ən çox tətbiq edilən xətaların geriyə yayılması (*error backpropagation*) alqoritmi bu prinsiplə işləyir. Supervizor öyrətmə alqoritminə sonra daha ətraflı baxılacaqdır.



Şəkil 6.5 Əks əlaqəsiz şəbəkə



Şəkil 6.6 Rekurent şəbəkə



Şəkil 6.7 Lateral ləngidici şəbəkə

Süni neyronların birlaylı şəbəkələri. Siqnalların paylanması üçün ayrılmış giriş neyronları layından başqa bir hesablayıcı neyron layı olan iyerarxik qeyri-rekurent şəbəkə sadə birlaylı şəbəkə adlanır. Hesablayıcı neyronlarda hər bir neyronun çıxış siqnalı onun girişinə daxil olan siqnalların uyğun əmsallara vurulduqdan sonrakı cəmindən asılı funksiya kimi təyin olunur. Ən sadə halda, çıxış hər bir hesablayıcı neyronun girişinə daxil olan siqnalların sadəcə olaraq uyğun əmsallara vurulduqdan sonrakı cəmidir. Çıxış siqnallarının toplusu şəbəkənin Y çıxış vektorunu əmələ gətirir. Y vektorununun m-ölçüsü şəbəkənin çıxışlarının sayına bərabərdir. Analoji qayda ilə n-ölçülü X giriş vektorunu, $n \cdot m$ ölçülü W çəki əmsalları matrisini təyin etsək, biz şəbəkənin çıxışının onun girişindən asılılığını vektor şəklində $\mathrm{Y} = \mathrm{XW}$ kimi yaza bilərik.

Göstərmək olar ki, fəallaşma funksiyası xətti olan belə şəbəkənin hesablama gücü yeni laylar əlavə ediləndə artmır. Birinci və ikinci laylarının çəki əmsalları matrisləri uyğun olaraq W_1 və W_2 olan ikilaylı şəbəkəyə baxaq. Birinci layın neyronlarının çıxış vektoru $Y_1 = XW_1$, ikinci layın çıxış vektoru isə $Y_2 = Y_1W_2 = XW_1W_2 = XW$ düsturu ilə tapılır, burada $W = W_1W_2$ verilmiş ikilaylı şəbəkəyə ekvivalent olan birlaylı şəbəkənin çəki əmsalları matrisidir.

Analoji qayda ilə xətti fəallaşma funksiyalı istənilən çoxlaylı şəbəkəni ekvivalent birlaylı şəbəkəyə çevirmək olar.

Biz yuxarıda şəbəkənin laylarının sayını göstərəndə paylaşdırıcı rolunu oynayan giriş layını nəzərə almamışıq. Bundan sonra layın sayını göstərəndə bütün layları nəzərə alacağıq.

Çoxlaylı süni neyron şəbəkələri. Qeyri-xətti fəallaşma funksiyasından istifadə edilən şəbəkədə layların sayının artırılması onun hesablama gücünün artırılmasına, daha mürəkkəb inikasların qurulmasına imkan verir.

Gizli layları olmayan sadə şəbəkələr bir sıra məsələləri həll edə bilmir. Belə məsələyə hamıya yaxşı məlum olan məşhur "istisna edidici və ya" (modul 2-yə görə toplama) problemi nümunə ola bilər ki, köməkçi laylar əlavə edilməyən sadə şəbəkəylə bu məsələni həll etmək mümkün deyil. Problemin məhz nədən ibarət olduğunu başa düşmək üçün neyronun funksiyasına daha müfəssəl baxaq. Tutaq ki, neyronun n girişi var. Hər bir koordinatı neyronun girişlərindən birinə uyğun gələn n-ölçülü koordinat fəzasına baxaq. Tutaq ki,

$$\sum_{i=1}^{n} w_i x_i - \theta = 0 \tag{6.3}$$

tənliyi verilmişdir.

Göründüyü kimi, (6.3) tənliyi n-ölçülü fəzada hipermüstəvini təyin edir. Bu hipermüstəvi fəzanı iki altfəzaya bölür. Bunlardan biri üçün $\sum_{i=1}^{n} w_i x_i - \theta > 0$ bərabərsizliyi, digəri üçün $\sum_{i=1}^{n} w_i x_i - \theta < 0$ bərabərsizliyi ödənəcəkdir.

Fərz edək ki, neyronun fəallaşma funksiyası

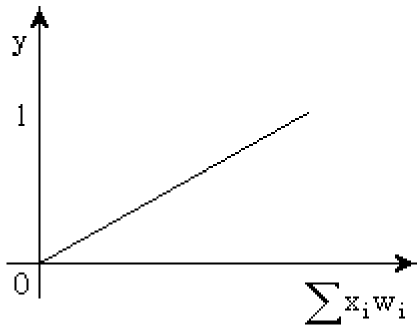
$$y = \begin{cases} 1, & \text{яэяр } \sum_{i=1}^{n} w_i x_i - \theta \geq 0 \\ 0, & \text{якс щалда} \end{cases}$$

şəklindədir. Onda belə nəticəyə gəlmək olar ki, neyron giriş siqnalları fəzasını iki hissəyə ayırır. Bunun sayəsində girişlərin müəyyən kombinasiyası neyronun fəallaşmasına səbəb olur, eyni zamanda başqa kombinasiyalar neyronu aktivləşdirmir. Buradan belə nəticəyə gəlmək olar ki, ikilaylı neyronu şəbəkə yalnız xətti separabel məsələləri həll edə bilər.

Determinik neyron şəbəkələr. Elementlərinin fəallaşma funksiyaları determinik olan neyron şəbəkələri determinik neyron şəbəkələri adlanır. Fəallaşma funksiyası kimi müxtəlif funksiyalardan istifadə olunur. Tədqiqatlarda ən çox istifadə olunan fəallaşma funksiyalarının əsas növlərini sadalayaq.

1. Xətti funksiya

$$y = k \sum_{i} x_i w_i$$



Şəkil 6.8 Xətti funksiya

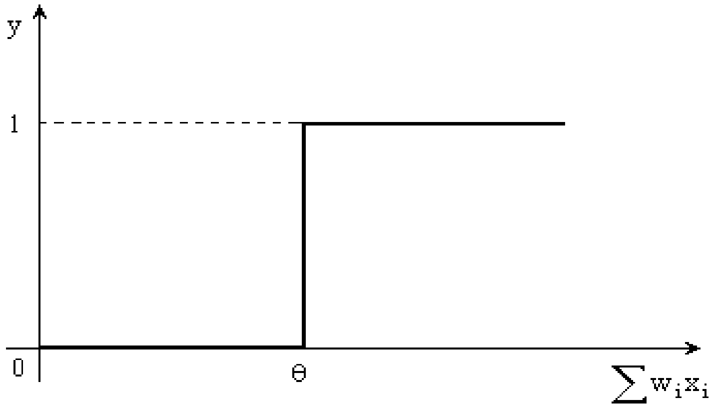
Xətti funksiyanın qrafiki şəkil 6.8-də verilmişdir. $k=1$ halında neyronun çıxışı onun giriş siqnallarının əmsallaşdırılmış cəmi olur. Bu funksiyanın tətbiq dairəsi olduqca məhduddur. İstənilən xətti elementli çoxlaylı şəbəkəni birlaylı şəbəkəyə gətirmək mümkün olduğundan bu funksiya əsasında çoxlaylı şəbəkənin qurulması onun hesablama gücünü artırmır.

2. Vahid sıçrayış funksiyası

$$y=\begin{cases}1, & \text{яэяр } \sum_{i=1} w_i x_i \geq \theta \\ 0, & \text{якс щалда}\end{cases}$$

burada θ hüdud qiymətidir.

Bu funksiyanın qrafiki şəkil 6.9-da gəstərilmişdir. Vahid sıçrayış funksiyasından istifadə edən neyronlar girişlərin əmsallara vurulmuş cəmi hüdud kəmiyyətinə çatanda ani olaraq özlərinin vəziyyətlərini “0”-dan “1”-ə dəyişir.

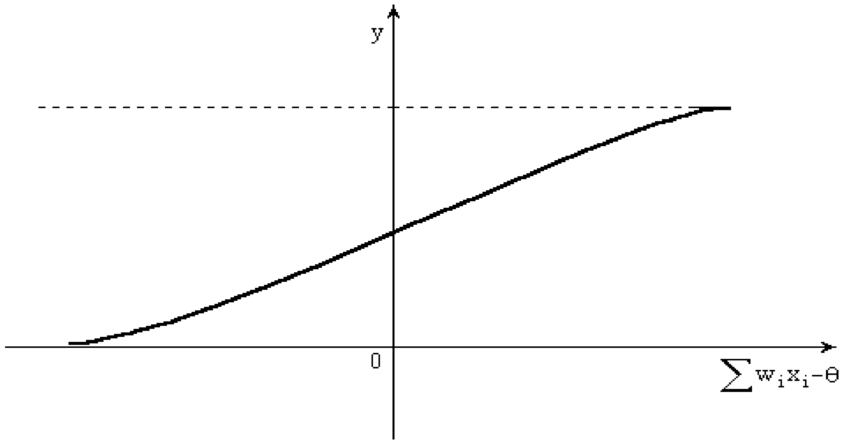


Şəkil 6.9 Vahid funksiya

3. Siqmoid (*yarımxətti funksiya*)

$$y=\frac{1}{1+e^{-\sum w_i x_i}}$$

Funksiyanın qrafiki şəkil 6.10-da verilmişdir. Bu fəallaşma funksiyasının verilməsinin ən çox yayılmış formasıdır. Ümumi giriş $-\infty$-a yaxınlaşdıqda fəallaşma səviyyəsi sıfıra yaxınlaşır, ümumi girişin çox böyük qiymətlərində fəallaşma səviyyəsi praktiki olaraq “1”-ə bərabərdir.



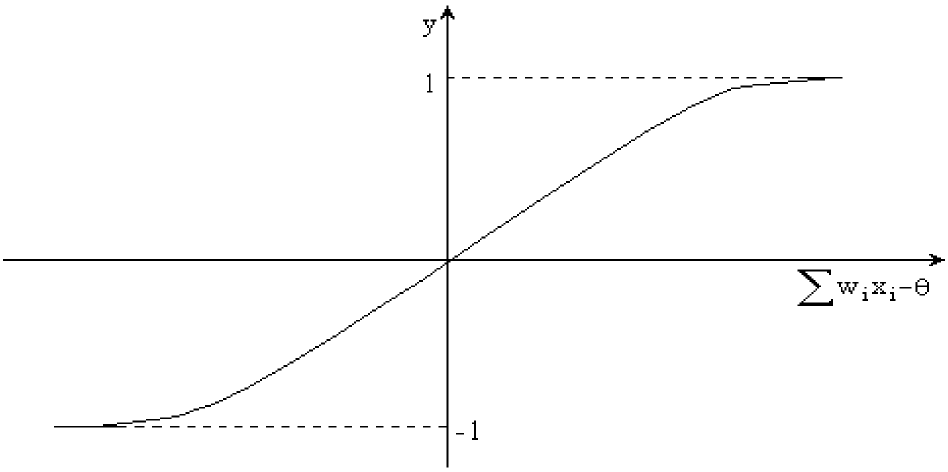
Şəkil 6.10 Siqmoid funksiyası

4. Hiperbolik tangens

$$y = th\left(\sum_i w_i x_i\right)$$

Şəkil 6.11-də bu funksiyanın qrafiki verilmişdir. Siqmoid funksiyasından fərqli olaraq, hiperbolik tangensdən istifadə ediləndə neyronların fəallaşma səviyyəsi –1-dən 1-ə qədər dəyişir.

Axırıncı üç funksiya mürəkkəb məsələləri həll etməyə imkan verən çoxlaylı neyron şəbəkələrinin qurulmasında xususi maraq kəsb edir. Bu funksiyaların qiymətlər oblastı ([0,1] və ya [-1,1]) çıxış vektorlarının təsnifatını aparmaq üçün geniş imkanlar açır. Artıq qeyd edildiyi kimi, xətti funksiya az istifadə olunur. Yuxarıda qeyd etdiyimiz kimi deter-minik neyron şəbəkələr həm əks əlaqəsiz (feed-forward), həm əksəlaqəli (*feed-back*) ola bilir. Əks əlaqəli neyron şəbəkələr daha böyük maraq kəsb edir. Qeyd edildiyi kimi, bu şəbəkələr rekurent şəbəkələr adlanır. Sadə və çoxlaylı qeyri-rekurent şəbəkələrdə giriş vektoru daxil olduqda şəbəkə üzrə giriş layından çıxış layına doğru istiqamətdə neyronların fəallaşması baş verir. Çıxış neyronlarının fəallaşma səviyyələri təyin edildikdən sonra neyronların vəziyyətlərində heç bir dəyişiklik aparılmır. Bu proses neyron şəbəkələrin relaksasiyası adlanır.



Şəkil 6.11 Hiperbolik tangens

Rekurent şəbəkələr özlərini tam başqa cür aparırlar. Şəbəkənin girişinə siqnal vektoru veriləndə neyronların vəziyyətləri müəyyənləşir, lakin sonra neyronların çıxışları əks əlaqəli olduqları üçün onların girişlərinə yenidən təzə vektor daxil olur və vəziyyətlər təzədən dəyişirlər. Rekurent şəbəkələrdə dayanıqlıq anlayışı çox vacibdir. Şəbəkə o zaman dayanıqlı hesab olunur ki, sonlu sayda iterasiyadan sonra neyronlar dəyişilməyən vəziyyət almış olsun. Dayanıqlı rekurent şəbəkələrin girişinə siqnal vektoru veriləndə neyronların çıxış siqnalları formalaşır ki, bu siqnallar sonra təzədən yeni vəziyyətlər vektoru generasiya etməklə bir də (təkrarən) girişlərə daxil olur. Lakin şəbəkənin son vəziyyəti qərarlaşmayana qədər iterasiyaların sayının artması ilə düyünlərin vəziyyətlərinin dəyişməsinin sayı azalır. Əks əlaqəsiz şəbəkələr həmişə dayanıqlıdırlar. Çünki, bir giriş vektoru üçün şəbəkənin düyünləri öz vəziyyətini yalnız bir dəfə dəyişə bilər.

Dayanıqsız şəbəkələr giriş vektorunu verəndə sonra təkrarən girişlərə ötürülən çıxış siqnalları yaranır və yeni vəziyyət dəyişmələrinə səbəb olur. Həm də bu proses sonsuz olaraq davam edir və qərarlaşmış bir vəziyyət yaranmır.

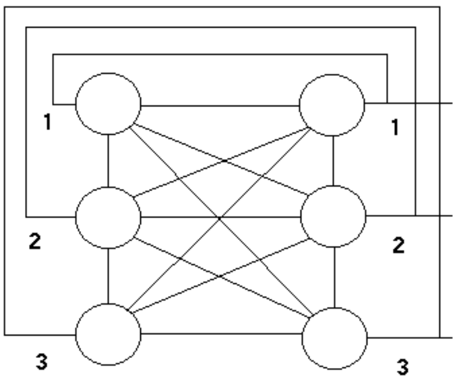
Dayanıqlı olmayan neyron şəbəkələrində həm də xaotik relaksasiya baş verə bilər ki, bu da tədqiqatçılar üçün böyük əhəmiyyət kəsb edir. Lakin xaotik relaksasiya problemlərinə bu kitabda baxılmır.

Neyron şəbəkələrin dayanıqlığını hansı meyarlarla təyin etmək məsələsi uzun müddət açıq qalmışdır. Lakin rekurent şəbəkələrin dayanıqlı alt çoxluğunu təyin etməyə imkan verən teoremin isbatı sayəsində [37] bu məsələ həll olundu və mürəkkəb rekurent şəbəkələrin tətbiqi üçün imkanlar artdı.

Neyron şəbəkələrin ən böyük tədqiqatçılarından biri olan Xopfild xüsüsi şəkilli şəbəkələr təklif etmişdir. Bu şəbəkələri Xopfild şəbəkələri adlandırırlar. Şəkil 6.12-də göstərilən sadə ikilaylı şəbəkəyə baxaq. Hər bir neyronun girişinə X giriş vektorunun uyğun komponentindən başqa birinci layın neyron-paylayıcıları vasitəsilə başqa neyronların çıxış siqnalları daxil olur. Xopfild özünün əvvəlki tədqiqatlarında [38] fəallaşma funksiyası olaraq

$$y_j = \begin{cases} 1, \text{ яэяр } \sum\limits_{i \neq j} \mathrm{w}_{\mathrm{ji}} y_i + x_j > \theta_j \\ 0, \text{ яэяр } \sum\limits_{i \neq j} \mathrm{w}_{\mathrm{ji}} y_i + x_j < \theta_j \\ \text{дяйишмир, яэяр } \sum\limits_{i \neq j} \mathrm{w}_{\mathrm{ji}} y_i + x_j = \theta_j \end{cases} \quad (6.4)$$

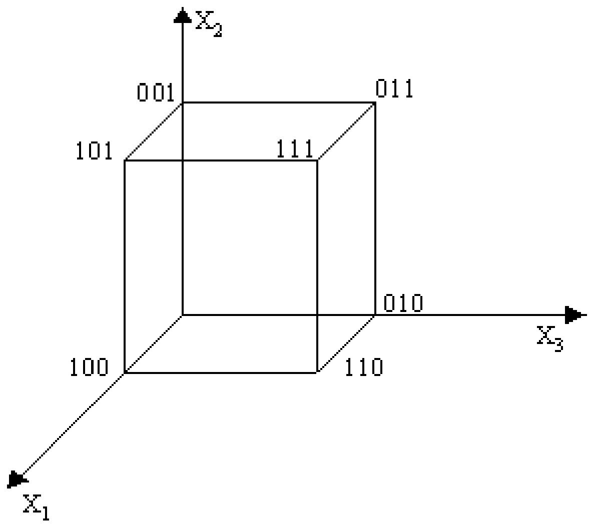
funksiyasını götürmüşdür.



Şəkil 6.12 Sadə rekurent şəbəkə

(6.4) ifadəsindən görünür ki, j-cu neyronun x_j girişi ilə digər neyronların əmsallaşdırılmış cəmi hüdudu aşsa, onda bu neyron fəal vəziyyətə keçir, hüdudu aşmasa passiv vəziyyətdə olur, cəm hüdud qiymətinə bərabər olduqda neyronun vəziyyəti dəyişmir. Xopfildin ilk işlərində neyronların hər birinin vəziyyəti diskret təsadüfi zaman anında, yəni asinxron dəyişilirdi. Çıxış siqnalları toplusu şəbəkənin ikili vəziyyət vektorunu formalaşdırır (yaradır).

Beləliklə, ikili qiymətlər alan n-çıxışlı neyron şəbəkə 2^n vəziyyətlərində ola bilər. Əgər n-ölçülü fəzada koordinat oxları üzərində neyronların fəallaşma səviyyələrinin qiymətlərini ayırsaq, bu vəziyyətlərin hər biri n-ölçülü fəzada hiperkubun bir təpəsini təmsil edəcəkdir. Giriş vektoru veriləndə şəbəkə bir vəziyyətdən digərinə keçir, yəni son vəziyyət alınanadək hiperkubun təpələri üzrə hərəkət edir. Dayanıqlı vəziyyət giriş vektorlarının, çəki əmsalları və hüdudların müəyyən qiymətlərinə uyğun gəlir. Belə ki, müəyyən bir şəbəkədə çəki əmsalları və hüdudların qiymətləri sabit qalırsa (öyrətmə rejiminə burada baxılmır), onda dayanıqlı vəziyyətlər cari giriş vektorları ilə tam müəyyən olunur. Şəkil 6.13-də üç çıxış neyronlu şəbəkə üçün vəziyyətlər kubu göstərilmişdir.



Şəkil 6.13 Üç çıxış neyronlu şəbəkə üçün vəziyyətlər kubu

Koxonen və Qrossberq [39] göstərmişlər ki, əgər rekurent şəbəkənin çəki əmsalları matrisi baş diaqonal elementləri 0 olan simmetrik matrisdirsə, onda bu şəbəkə dayanıqlıdır. Bu hökm aşağıdakı kimi isbat olunur. Tutaq ki, şəbəkənin vəziyyətinin hər bir dəyişməsində öz qiymətini azaldan funksiya var. Aydındır ki, bu funksiya sıfıra çatanadək azalmalıdır. Bu xassəni ödəyən funksiya olaraq Lyapunov funksiyasını götürmək olar

$$E = -\frac{1}{2}\left(\sum_i \sum_j w_{ij} y_i y_j - \sum_j x_i y_j + \sum_j \theta_j y_j \right). \tag{6.5}$$

Burada E şəbəkənin "enerjisi" adlanır: x_j kənardan verilən giriş vektorunun j-cu komponenti, θ_j neyronun hüdududur.

(6.5) ifadəsini nəzərə alaraq j-cu neyronunun vəziyyətinin dəyişməsindən yaranmış enerji dəyişməsini

$$\Delta E = -\left[\sum_{i \neq j} \left(W_{ij} Y_i \right) + X_j - \theta_j \right] \Delta y_j \tag{6.6}$$

düsturu ilə ifadə etmək olar.

(6.6) ifadəsinə baxaq. Əgər neyronun ümumi ölçülmüş girişi hüdud kəmiyyətini aşırsa, onda o aktiv vəziyyətə keçir. Bu isə onu göstərir ki, kvadrat mötərizənin içərisində olan ifadənin müsbət qiymətində y_j "1" qiymətini alır. Beləliklə, əgər buna qədər onun qiyməti 0 idisə, onda artım kəmiyyəti $\Delta y_j = 1$, əgər neyron buna qədər də aktiv idisə, onda onun vəziyyəti dəyişmir və $\Delta y_j = 0$. Bu onu göstərir ki, ΔE kəmiyyəti ya mənfidir, ya da sıfırdır.

Əgər ümumi giriş hüdud qiymətindən kiçikdirsə, neyron “0” vəziyyətini alır. Bu da onu gəstərir ki, Δy_j kəmiyyəti ya “0”, ya da “-1”-dir. Bütün (6.6) ifadəsi bu halda ya mənfi, ya da sıfıra bərabər olacaqdır.

Ümumi girişin hüdud qiymətinə bərabər olduğu halda (6.6) ifadəsinin hər iki vuruğu sıfır olacaqdır, deməli, ΔE də sıfıra bərabər olacaqdır.

Beləliklə, neyronların vəziyyətlərinin istənilən dəyişməsində şəbəkənin enerjisi ya azalacaq, ya dəyişməyəcək və vəziyyətlərin sonlu sayda dəyişməsindən sonra o sıfra yığılacaq və sonra dəyişməz qalacaqdır ki, bu da şəbəkənin dayanıqlıq əlamətidir.

Beləliklə, simmetrik çəki matrisli şəbəkələrdə onun davranışı dayanıqlıdır. Qeyd etmək lazımdır ki, çəki matrisinin simmetrik olması şəbəkənin dayanıqlı olması üçün kafi şərtdir, amma zəruri şərt deyil.

Rekurent şəbəkələr assosiativ yaddaş modellərinə nümunə ola bilər. İnsanın yaddaşı ilə ənənəvi kompüter yaddaşı arasında köklü fərq var. Maşın yaddaşından informasiyanı oxumaq üçün onun yerləşdiyi ünvanı göstərmək tələb olunur. İnsan yaddaşı bir qədər başqa cür təşkil edilmişdir. İnsan yaddaşından informasiya götürmək üçün məzmuna uyğun bəzi açar informasiya tələb olunur. Məsələn, əgər adam ona tanış olan kitabın adını eşidirsə, o, kitabın müəlliflərini, məzmununu, cildinin rəngini və s. yada sala bilər. Poeziya biliciləri şerin birinci sətirlərinə görə bütün şeri tamamilə yadına sala bilərlər, musiqi biliciləri fraqmentinə görə bütün əsəri yaddaşında bərpa edə bilərlər və s.

Rekurent şəbəkələrdə assosiativ yaddaşı modelləşdirmək üçün onların məqsəd vəziyyətlərini müəyyən etmək və vəziyyətlər fəzasının bu nöqtələrində enerjinin minimallaşdırılmasını təmin etmək lazımdır. Əgər bu şərtlər ödənilərsə, onda tam surətin hissəsi veriləndə və ya xarab olmuş variantında şəbəkə yüksək ehtimalla tam surəti canlandıra biləcəkdir. Rekurent şəbəkələrin verdiyi bu imkan informasiyanın bərpa olunması üsullarının işlənməsində əhəmiyyətli rol oynaya bilər.

Xopfild [39]-da assosiativ yaddaşın işlənməsi üçün aşağıdakı ideyanı təklif etmişdir. Tutaq ki, hər biri binar vektor olan $Y^1, Y^2, \ldots, Y^m$ surətlərini yadda saxlamaq lazımdır. Onda rekurent şəbəkələr əsasında bu surətləri “yadda saxlayan” assosiativ yaddaş yaratmaq üçün onun çəki əmsallarını $w_{ij} = \sum_{k=1}^{m} y_i^k y_j^k$ düsturu ilə hesablamaq lazımdır, burada y_i^k k-cı vektorun i-ci komponentidir.

Beləliklə, çəki əmsallarını təyin etdikdən sonra (bu halda $w_{ij} = w_{ji}$ olması şəbəkənin dayanıqlığını təmin edir) şəbəkə assosiativ yaddaş kimi istifadə oluna bilər. Əgər indi onun girişinə y^j-nin natamam və ya korlanmış variantı kimi təqdim edilən y^{j*} vektoru verilsə, şəbəkə çox yüksək ehtimalla orijinalı bərpa edəcəkdir. Bu ona görə belə olur ki, giriş vektoru veriləndən sonra şəbəkə “zəifləməyə” başlayacaqdır, yəni tədricən öz enerjisini yaxınlıqdakı minimuma tərəf azaldacaqdır. Bu minimumun qlobal deyil, lokal olduğu variant da istisna edilmir.

Bu vaxta qədər biz elementləri vahid fəallaşma funksiyasına malik olan rekurent şəbəkələrə baxırdıq. Siqmoid funksiyası bioloji həqiqətə vahid funksiyadan daha uyğundur, belə ki, neyronun bir vəziyyətdən başqa vəziyyətə keçməsi üçün müəyyən müddət tələb olunur. Xopfild fəallaşma funksiyası

$$y = 1 \Big/ \left(1 + e^{-\lambda \sum x_i w_i}\right) \tag{6.7}$$

şəkilli siqmoid funksiya şəbəkələri də araşdırmışdır [4].

Burada λ əmsalı funksiyanın dikliyini göstərir. λ sonsuzluğa yaxınlaşanda (6.7) funksiyası vahid funksiyaya yaxınlaşır, λ kiçiləndə əyri hamarlanır.

Bu şəbəkələr üçün də dayanıqlığın kafi şərti, yəni çəki əmsalları matrisinin simmetrikliyi ödənilir.

Stoxastik neyron şəbəkələri. Real aləmdə tez-tez stoxastik proseslərə rast gəlinir. Real sistemin davranışına təsir edən bütün amillərin nəzərə alınması çətinliyi və bəzən də mümkün olmaması belə sistemlərə stoxastik sistemlər kimi baxmağa məcbur edir. Ən məşhur stoxastik neyron şəbəkələri sinfi Bolsman maşınları kimi tanınan sinifdir [4,5,6]. Onlar Xopfild şəbəkələri prinsipi əsasında qurulurlar, yəni onlarda simmetrik çəki əmsalları matrisindən istifadə olunur, lakin burada neyronların fəallaşma qaydası determinik deyil, stoxastikdir. Bu onu göstərir ki, verilmiş giriş siqnallarında neyronun çıxış siqnalının səviyyəsini bilavasitə təyin edən funksiya əvəzinə burada neyronun fəallaşması ehtimalını müəyyənləşdirən münasibət istifadə olunur.

6.2-nin əvvəlində Xopfild şəbəkələrinin təsviri verilmişdir. Bu şəbəkələrlə enerji və stabillik anlayışları əlaədardır. Şəbəkənin enerjisini təyin edən aşağıdakı münasibətə baxaq.

$$E = \sum_{i<j} w_{ij} y_i y_j + \sum_i \theta_i y_i \tag{6.8}$$

Burada y_i və y_j uyğun olaraq i-ci və j-cu neyronların aktivlik səviyyələri; θ_i i-ci neyronun hüdududur. k-cı neyronun fəallaşması şəbəkənin ümumi enerjisinin

$$\Delta E_k = \sum_i w_{ki} y_i - \theta_k \tag{6.9}$$

kəmiyyəti qədər azalmasına səbəb olur.

Determinik Xopfild şəbəkələrində bu faktdan assosiativ yaddaşın qurulmasında istifadə etmək olar. Bu halda öyrənmə çoxluğunun hər bir xüsusi vektoru üçün (yəni hər ayrıca yaddaş üçün) E enerjisinin minimuma endirilməsi təmin olunur. Yuxarıda qeyd edildiyi kimi, şəbəkənin girişinə küylü və yaxud da natamam xarici vektor daxil olarsa, belə şəbəkə lokal minimuma çatmayana və dayanıqlı vəziyyət almayana qədər öz enerjisini tədricən azalda bilər. Göstərilən yanaşma qlobal minimumun alınmasına zəmanət vermir. Buna görə də Erkli və onun əməkdaşları [5] belə sistemlərin davranışına təsadüfilik elementi daxil etmişlər.

(6.9) münasibətindən görünür ki, ΔE_k kəmiyyəti şəbəkənin ümumi çəkilərə vurulmuş girişlərinin cəmindən başqa bir şey deyildir. Stoxastik şəbəkələrdə fəallaşma qaydası ehtimalların paylanmasının f sıxlıq funksiyası ilə verilir:

$$p_k = f(\Delta E_k) = f\left(\sum_i w_{ki} y_i - \theta_k\right),$$

burada p_k k -cı neyronun aktivləşmə ehtimalıdır.

f funksiyası kimi vahid funksiya istifadə edilə bilməz, çünki bu halda şəbəkə determinik olur. Xətti funksiyadan və hiperbolik tangensdən istifadə etmək münasib deyil, çünki onların qiymətlər oblastı [0,1] parçasından kənara çıxır, ehtimalın qiymətləri isə [0,1] parçasında yerləşir. Buna görə də ehtimalların paylanma sıxlığı funksiyası kimi çox vaxt

$$P_k = \frac{1}{1 + e^{-\Delta E_k / T}}, \tag{6.10}$$

siqmoid funksiyadan istifadə olunur. Burada T şəbəkədə təsadüfilik səviyyəsini təyin edən hər hansı parametrdir. T parametri “temperatur” adlanır, bu ondan irəli gəlmişdir ki, Bolsman maşını ilə termodinamik sistem arasında paralellik aparmaq olar.

Metalların tavlanması prosesinə baxaq. Metallarda atomlar aşağı temperaturlarda kristal qəfəsin düyünlərində yerləşirlər. Onlar möhkəm əlaqələnmişlər və onların enerjisi qəfəsin müəyyən etdiyi mövqeni tərk etməyə imkan vermir. Metal qızdırılanda atomların enerjisi artır və onların qəfəsin düyünlərinin müəyyən etdiyi mövqe ətrafında rəqsinin amplitudu artır. Temperatur ərimə nöqtəsindən yuxarı çatdıqda qəfəsin quruluşu pozulur və atomlar yüksək enerjiyə malik olmaqla sərbəst hərəkət etməyə başlayırlar. Temperatur aşağı düşəndə atomlar yenidən enerji itirərək enerjinin minimumuna uyğun düzülüş təşkil edirlər.

Oxşar proses Bolsman maşınlarında imitasiya olunurlar, ona görə də o “imititasiya olunan tavlama” (simulated annealing) adlanır. Şəbəkədə neyronların vəziyyətlərinin təzələnməsinin lazımi qədər böyük sayında, yəni (6.10) funksiyası iterativ tətbiq ediləndə şəbəkənin iki qlobal A və B vəziyyətlərində (burada qlobal vəziyyət dedikdə şəbəkənin elementlərinin vəziyyətləri toplusu nəzərdə tutulur) olması ehtimallarının nisbəti sabit qalır:

$$\frac{P_A}{P_B} = e^{-(E_A - E_B)/T}$$

Belə halda deyirlər ki, şəbəkə “temperatur tarazlığı”na çatmışdır. Aydındır ki, temperatur tarazlığında şəbəkənin elementləri öz vəziyyətlərini dəyişə də bilərlər, belə ki, fəallaşma funksiyası – ehtimalların paylanma sıxlığı 0 və 1 qiymətini yalnız ümumi giriş uyğun olaraq -∞ və ∞-ğa yaxınlaşanda ala bilir, yəni $\lim\limits_{\Delta E_k \to -\infty} p_k = 0$, $\lim\limits_{\Delta E_k \to +\infty} p_k = 1$

Şəbəkənin hər hansı bir qlobal A vəziyyətində olması

$$P_A = e^{-E_A / T}, \tag{6.11}$$

düsturu ilə verilir. Bu Bolsman paylanmasıdır. Elə buradan da “Bolsman maşını” adı götürülmüşdür. (6.11) münasibətindən görünür ki, yüksək temperaturlarda şəbəkənin aşağı enerjili səviyyədə olması ehtimalı aşağıdır, bu halda şəbəkə yəqin ki, yüksək enerjili vəziyyətlərdə ola bilər. Temperatur aşağı düşəndə əks tendensiya müşahidə olunur: aşağı enerjili vəziyyətlər yüksək enerjili vəziyyətlərə nisbətən çox ehtimallıdır. Belə görünə bilər ki, şəbəkənin az enerjili vəziyyətə çatması üçün aşağı temperaturlar tətbiq olunmalıdır (xatırladaq ki, məsələ enerjinin minimuma endirilməsi məsələsidir). Lakin aşağı temperaturlarda şəbəkə temperatur tarazlığına çox yavaş çatır. Digər tərəfdən, yüksək temperaturlarda temperatur tarazlığı tez alınır.

Yuxarıda göstərilən problemin kompromis həlli əvvəlcə yüksək temperaturun verilməsi, sonra isə onun tədricən aşağı salınmasıdır. Beləliklə, şəbəkə elə bil ki, metallar kimi tavlanır. "İmitasiya olunan tavlama" şəbəkəyə nisbətən az sayda iterasiyada aşağı enerjili vəziyyətə çatmağa imkan verir.

Çox böyük olmayan məsələlər üçün şəbəkə kifayət qədər münasib həll təklif edə bilər (yəni giriş vektorunu verəndə kifayət qədər yüksək dəqiqliklə uyğun çıxış vektoru generasiya olunacaqdır). Lakin daha böyük məsələlər üçün sadə birlaylı stoxastik şəbəkələrin istifadəsi çətinlik törədir. Bu çətinlik şəbəkənin hesablama gücünün və deməli, məsələnin həllinin dəqiqliyinin artırılmasını təmin edən əlavə gizli layların daxil edilməsi hesabına həll edilə bilər.

Qeyri-səlis neyron şəbəkələr. İndiyə kimi baxdığımız adi və ya "səlis" neyronları ümumi şəkildə aşağıdakı kimi təsvir etmək olar. Süni neyronlar – bunlar bir neçə girişi ($n, n > 1$) və bir çıxışı olan prosessor elementləridir. Hər bir girişə müəyyən $f_i(x_i)$ funksiyası uyğun tutulur ki, bu funksiya giriş siqnalını müvafiq qaydada emal edir. Ən sadə və ən çox yayılmış halda bu funksiya giriş qiymətini əlaqələrin çəki əmsallarına vurmaqla onu gücləndirir, yəni $f_i(x_i) = x_i \cdot w_i$. Neyronun çıxış qiyməti isə giriş parmetrləri f_i funksiyalarının verdiyi qiymətlər olan n-ölçülü $g(y_1, y_2, \ldots, y_n)$ funksiyasıdır. g funksiyası kimi çox hallarda $g(I) = 1/(1+e^{-I})$ siqmoid funksiyası götürülür. Burada I f_i funksiyalarının bütün əlaqələr üçün qiymətləri və müəyyən bir θ sabitinin cəmidir

$$I = \sum_i f_i(x_i) + \theta$$

θ tərs hüdud və ya sürüşmə (*bias*) adlanır.

Qeyri-səlis neyronda yeni anlayışlar – qeyri-səlis ədəd, mənsubiyyət funksiyası, qeyri-səlis əməliyyatlar anlayışları istifadə olunur. Bu halda f və g qeyri-səlis dəyişənli və qeyri-səlis qiymətli funksiyalar ola bilər.

Neyron şəbəkələrin bəzi xüsusiyyətlərini təhlil edək. İnsan beyni sinaptik əlaqələr vasitəsi ilə bir şəbəkədə birləşmiş 14 milyarda yaxın neyrondan ibarətdir. İnsanda qavrama, duyğu, təfəkkür proseslərinin hamısı bu şəbəkədə elektrik siqnallarının keçməsi hesabına həyata keçirilir. Əvvəlki paraqrafda təsvir olunmuş bioloji neyronlarda siqnalların real emalı prosesi çox mürəkkəbdir və kifayət qədər öyrənilməmişdir. Neyronların işə düşməsi real vaxt şəraitində asinxron baş verir. Neyronların fəallığı zamana görə dəyişir. Sinaptik əlaqələrin gücü də dəyişir, amma çox yavaş. Biz daha tez fikirləşirik, nəinki öyrənirik [40]. Əgər i-ci neyronun vəziyyətini x_i ilə, əlaqənin j-cu çəki əmsalını isə w_j ilə işarə etsək, onda

$$\frac{dw_j}{dt} << \frac{dx_i}{dt}$$

yazmaq olar.

Neyronların aktivliyinin dəyişməsi neyrona daxil olan siqnalların qiymətindən, həm də onların təsir müddətindən (sürəkliliyindən) asılıdır. Əlaqələrin çəkilərinin dinamikası xüsusi maraq kəsb edir. Xebbin [41] çəki əmsallarının dəyişilməsi haqqında söylədiyi ideya bioloji həqiqətə uyğun sayılır. O qeyd etmişdir ki, əgər bir-biri ilə əlaqəli iki neyron eyni zamanda fəallaşırsa, bir neyronun aksonunu digərinin gövdəsi (və ya dendriti) ilə əlaqələndirən düyünlərinin sayı artır, ya da bu düyünlərin ölçüləri böyüyür. Bu və digər hal iki neyron arasında əmsallaşdırılmış əlaqənin gücləndirilməsi deməkdir. Xebbin öyrətmə alqoritmi bu ideyaya əsaslanmışdır. Bu neyron şəbəkələrdə öyrətmə alqoritminin tarixən birincisidir.

Süni neyron şəbəkələri müəyyən quruluşda birləşmiş süni neyronlar çoxluğundan ibarətdir. Neyroşəbəkə arxitekturalarının müxtəlif tipləri 6.2-də təsvir edilmişdir.

İndi isə neyron şəbəkələrin bəzi xassələrinə baxaq.

Neyron şəbəkələrin əsas xassələrindən biri şəbəkələrdə informasiyanın paralel emalıdır. Neyronların hər biri başqa elementlərin hər hansı qrupu ilə paralel olaraq işləyən ayrıca hesablama qurğusudur. Hər bir neyron şəbəkənin çıxış siqnalları vektorunun formalaşdırılmasında öz hesablama payı ilə iştirak edir. Buna görə də paralel aparatlarda onların realizə olunması neyron hesablamaların yüksək sürətini müəyyənləşdirir.

Paralellik prinsipindən neyron şəbəkələrində informasiyanın paylanmış təsviri irəli çıxır. Əgər kiməsə “Bir əlin barmaqları ilə neçə ədədi göstərmək olar” sualını versək, onda yəqin ki, o “beş” deyə cavab verəcəkdir. Lakin bükülü və açıq barmaqların kombinasiyalarının köməyi ilə 32 ədədi göstərmək olar. İş burasındadır ki, birinci halda 1-lik (bir əsaslı) say sistemi, ikincidə isə ikilik say sistemi istifadə olunur. 1-lik say sistemində lokal təsvir istifadə olunur, əsası n ($n > 1$) olan sistemlərdə paylanmış təsvirdən istifadə olunur [2]. Paylanmış təsvir zamanı hər bir element hər bir obrazın təsvirində iştirak edir.

Neyroşəbəkə modellərinin etibarlılığı daha əhəmiyyətli bir xassədir. Burada etibarlılığı iki cür başa düşmək olar. Bir tərəfdən şəbəkənin girişinə küylü və ya natamam obraz daxil olduqda şəbəkə özünü obrazın orijinal variantı daxil olduğu kimi apara bilər. Digər tərəfdən hər hansı bir əlaqənin və ya tam neyronun sıradan çıxması ilə şəbəkənin işinin keyfiyyəti hiss olunmaz dərəcədə dəyişir. Bu xassə yüksək etibarlılıq tələb olunan sahələr üçün tətbiqi neyroşəbəkə sistemlərinin qurulmasına imkan verir.

Neyron şəbəkələrin əhəmiyyətli xassələrindən biri də onların öyrənmə qabiliyyətidir. Şəbəkə tək giriş vektorları çoxluğunu inikas edən avtomat funksiyanı yerinə yetirmir, o həmçinin öz parametrlərini (quruluşlarını, çəki əmsallarını) modifikasiya edir, beləliklə konkret məsələnin tələblərinə uyğun olaraq öz işini uyğunlaşdırır. Məsələn, supervizor alqoritmlərində giriş vektorları çoxluğunun çıxış vektorları çoxluğuna inikas etdirmək tələb olunur. Bu zaman şəbəkənin parametrləri konkret alqoritmdən asılı olaraq seçilir. Bu xüsusiyyət şəbəkənin parametrlərini konkret məsələ haqqında əhəmiyyətli biliyə malik olmadan hesablamağa imkan verir. Məsələn, neyron ekspert sistemlərin qurulması zamanı kifayət qədər statistik məlumat olarsa, ekspertlərlə işləmək kimi çox zəhmət tələb edən proses aradan çıxır (ekspert sistemlərinin qurulmasındakı problemlərlə tanış olanlar bilirlər ki, onların işlənməsində ən çətin biliklərin əldə olunmasıdır).

Neyron şəbəkələrin ümumiləşdirmə qabiliyyəti də onun ən mühüm xüsusiyyətlərindəndir. Bu xüsusiyyət sayəsində şəbəkələr təkcə öyrənmə vaxtı verilən inikası təmin etmir, həm də yenilərini qurur. Bu, neyron şəbəkələrə əsaslanan sistemlərin "səlahiyyətliliyini" artırır və onlara funksiyanın universal aproksimatoru rolunu oynamağa imkan verir.

Neyron şəbəkələrlə həll edilən məsələlər. Neyron şəbəkələrin tətbiqinin daha səmərəli olduğu bəzi məsələlərə baxaq. Bu məsələləri şərti olaraq aşağıdakı kimi ayırmaq olar:

– təsnifat;

– klasterləşdirmə;

– aproksimasiya;

– avtoassosiasiya .

1) Təsnifat (*klassifikasiya*). Təsnifat məsələsində obyektin $X^n = \{x_1, x_2, \ldots, x_n\}$ əlamətləri vektoru verilir və, obyektin m sayda kəsişməyən C_i siniflərindən birinə aid edilməsi tələb olunur:

$$C_i \cap C_j = \varnothing, \quad i \neq j\ ,\ i, j = \overline{1, m}\ .$$

Məsələn, uçan obyektlərin təsnifatında qanadların, mühərrikin və s. mövcudluğu əlamətlər ola bilər. Burada siniflər Təyyarə, Quş, Raket, Uçan Naməlum Obyekt və s.ola bilər. Əlamətlərin konkret qiymətlərinin hər bir kombinasiyası vektor verir. Bu vektor təsnif olunan sistemin girişinə daxil olan zaman sistem obyektin hansı sinifə aid olduğunu göstərməlidir.

Neyron şəbəkələri üçün dediklərimiz aşağıdakı kimi interpretasiya olunur. n giriş və m çıxış neyronları olan elə şəbəkə qurulmalıdır ki, girişə əlamətlər vektoru daxil olarkən ən yüksək fəallaşma səviyyəsinə malik olan çıxış elementi obyektin sinfini göstərsin. Bu imkanı təmin etmək üçün uyğun alqoritm seçməklə şəbəkənin öyrədilməsi aparılır.

2) Klasterləşdirmə. Klaster analizi məsələsində giriş əlamətlər vektorları çoxluğunu klasterlərə (oxşarlıq dərəcəsinə görə təşkil edilən obyektlər qrupu) elə ayırmaq tələb olunur ki, bir klasterə aid olan vektorlar bir-birinə yaxın, müxtəlif klasterlərdən olan vektorlar bir-birindən uzaq olsun.

Bu məsələni həll etmək üçün giriş neyronlarının sayı əlamətlər vektorunun komponentlərinin sayına, çıxış elementləri isə klasterlərin sayına bərabər olan şəbəkə qurulur. Çıxış neyronlarının çəki əmsalları vektoru uyğun klasterlərin mərkəzinə yaxın olan əlamətlər öyrətmə proseduru ilə seçilir.

6.2-də izah olunan Rəqabətli öyrətmə alqoritmi bu məsələni həll etməyə imkan verir.

3) Aproksimasiya. Neyron şəbəkələr funksiyaların universal aproksimatoru vəzifəsini yerinə yetirə bilər. Fərz edək ki, hər hansı $F(x)$ funksiyası verilib. Aproksimasiya məsələsi, istənilən kiçik $\varepsilon \geq 0$ üçün elə $F^*(x)$ funksiyasının seçilməsindən ibarətdir ki, $d\left(F(x), F^*(x)\right) < \varepsilon$ şərti ödənsin, burada $d(\ )$ - funksiyalar arasında məsafədir.

Ümumi halda aproksimasiya olunacaq $F(\ )$ funksiyasının şəkli bizə məlum deyil. Bizə yalnız $x_1 \to y_1, \ldots, x_n \to y_n$ şəkilli cütlər çoxluğu məlumdur, burada x_i funksiyanın kiriş dəyişənləri vektorun i -ci qiyməti, y_i funksiyanın i -ci qiymətidir. Klassik üsullardan istifadə etdikdə (məsələn, ən kiçik kvadratlar üsulundan) bizə aproksimasiya edən funksiyanın modelini (xətti, kvadratik və s.) seçmək və sonra seçilmiş modelin parametrlərini qiymətləndirilmək (identifikasiya) lazımdır.

Neyron şəbəkələr universal aproksimatordur, yəni kifayət miqdarda neyronlar veriləndə, onlar istənilən funksiyanı istənilən ε dəqiqliyi ilə aproksimasiya edə bilər [40]. Həm də qeyd etmək lazımdır ki, modelin şəklinin seçilməsinə zərurət qalmır.

Neyron aproksimasiyasında şəbəkənin öyrədilməsi üçün funksiyanın dəyişənlərinin və qiymətlərinin müşahidə cütləri istifadə olunur. Öyrənmədən sonra şəbəkə verilən inikası yüksək dəqiqliklə yerinə yetirir. Dəyişənlərin kombinasiyası N -ölçülü fazada nöqtəni verir, burada N -funksiyanın dəyişənlər vektorunun ölçüsüdür. Bu nöqtələrdən hər biri funksiyanın qiymətlərinin bir ölçülü fəzasında bir nöqtə ilə bağlıdır. Hər hansı x_i -nin ətrafında bir nöqtə seçilərkən şəbəkə $\mathrm{y_i}$ nöqtəsi ətrafında bir nöqtə təyin edir (neyron şəbəkələrin ümümiləşmə qabiliyyəti sayəsində), belə ki, bu ətraf nə qədər kiçik olsa, aproksimasiya bir o qədər dəqiq olur. Beləliklə, (X, Y) cütləri nə qədər çox olsa, aproksimasiya bir o qədər dəqiq olar.

Praktikada neyron aproksimasiyası dəqiq riyazi funksional modelin qurulması çətin olan obyektlərin identifikasiyasını aparmağa imkan verir.

4) Avtoassosiasiya. Belə məsələlərin həlli assosiativ yaddaş modellərinin yaradılması ilə əlaqədardır.

Kompüterin yaddaşı və insan yaddaşı müxtəlif prinsiplərlə təşkil olunur. Bizə lazım olan informasiyanı əldə etmək üçün kompüterin yaddaşında bu informasiyanın yazıldığı ünvanı bilmək lazımdır.

Məzmununa görə ünvanlanan yaddaş və ya assosiativ yaddaş adlanan başqa bir model də mövcuddur. Bu yaddaşdan informasiya əldə etmək üçün yadda saxlanınan obrazın bir hissəsini və ya onun xarab olmuş bir variantını vermək kifayətdir. Bu zaman axtarılan obrazın düzgün variantı meydana çıxır.

Assosiativ yaddaşlı neyron modelində obrazların yadda saxlanması uyğun prosedurların köməyi ilə neyronlar qrupu vasitəsilə təmin olunur. Şəbəkənin girişinə obrazın bir hissəsi daxil olarsa, şəbəkə korrekt tam obrazı verəcək. Obrazın əldə edilməsi müddəti yadda saxlanılan obrazların sayından asılı deyil [40].

6.2. Neyron Şəbəkələrin Öyrətmə Alqoritmləri

Öyrətmə üsullarının təsnifatı. Neyron şəbəkələri çox vacib öyrətmə xassəsinə malikdir. Bu xassə sayəsində onlar nəinki girişə daxil olan obrazı tanıya bilir, həm də münasib prosedurların köməyi ilə tanıma dəqiqliyini artırmaq üçün sazlana bilir. Beləliklə, neyron şəbəkələr iki rejimdə, öyrətmə və tanıma rejimlərində fəaliyyət göstərə bilərlər.

Neyron şəbəkələr qurulan zaman çəki əmsalları və hüdud qiymətləri təsadüfi verilir. Bu başlanğıc vəziyyətində şəbəkə hələ tanıma qabiliyyətinə malik deyil. Bu imkanı təmin etmək üçün məsələnin xarakterindən asılı olaraq bu və ya digər öyrətmə proseduru tətbiq olunur. Şəbəkənin lazımı dəqiqliklə tanıma aparması üçün öyrətmə xeyli maşın vaxtı tələb edir. Bu prosedurun gedişində neyron şəbəkənin parametrləri (çəkiləri, hüdudları, bəzi alqoritmlərdə isə şəbəkənin özünün quruluşu) tələb olunan tanıma dəqiqliyini təmin etmək üçün modifikasiya edilir. Bəzən bu proses arzuolunan nəticəni vermir. Belə hallarda seçilmiş öyrətmə prosedurunu çəki əmsallarının başqa başlanğıc qiymətlərində və ola bilsin ki, prosedurun özünün parametrlərinin qiymətlərini dəyişməklə təkrar etmək lazım kəlir.

Hazırda dünyada neyron şəbəkələrin çoxsaylı öyrətmə alqoritmləri mövcuddur. Tədqiqatçılar öyrətmə alqoritmlərinin təsnifatının bir neçə variantını təklif edirlər. Məsələn, Rummelhart və Mak-Klelland [7] həll edilən məsələrə görə aşağıdakı sinifləri ayırmaqla təsnifat vermişlər:

– obrazların assosiasiyası;
– avtoassosiasiya ;
– qanunauyğunluqların aşkar edilməsi;
– təsdiq etməklə öyrətmə.

Təsnifatı öyrətmə qaydalarının növlərinə körə də aparmaq olar. Bu zaman aşağıdakı növləri ayırmaq olar:

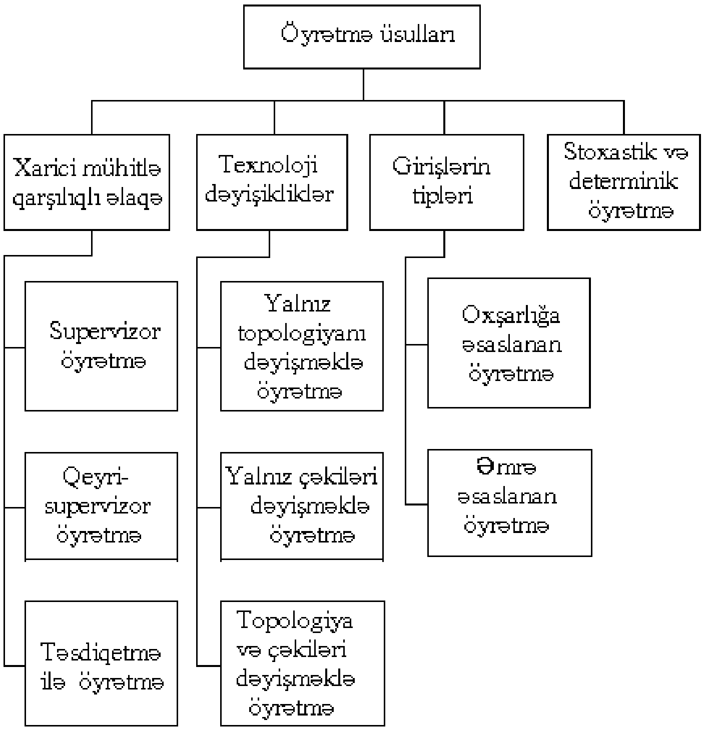
– korrelyasiya öyrətməsi;
– xətaları bilavasitə korreksiya etməklə öyrətmə;
– rəqabətli öyrətmə;
– Bolsman maşını vasitəsilə öyrətmə.

Şəkil 6.14-də [3]-də təklif olunmuş təsnifat verilmişdir. Bununla ətraflı tanış olaq.

Şəkildən göründüyü kimi, təsnifat aparmağın əlamətlərindən biri öyrədilən şəbəkənin xarici mühitlə (müəllimlə) qarşılıqlı əlaqənin xüsusiyyətidir.

Təsnifat üsüllarının müxtəlifliklərinə baxmayaraq onlar arasında öyrədilən şəbəkənin xarici mühitlə qarşılıqlı əlaqə məsələsi mühüm əhəmiyyət kəsb edir. Burada öyrətmə alqoritmləri öyrətmə zamanı xarici mühitdən daxil olan informasiyanın miqdarı və semantikasına uyğun olaraq müxtəlif siniflərə ayrılır. Bu baxımdan aşağıdakı kimi müxtəlif alqoritmlər sinfi mövcuddur:

– supervizor öyrətmə (supervised learning)
– qeyri-supervizor öyrətmə (unsupervised learning)
– təsdiqləmə ilə öyrətmə (reinforcement learning)



Şəkil 6.14 Neyron şəbəkələrin öyrətmə üsullarının təsnifatı

Xarici mühitlə ən sıx qarşılıqlı əlaqə supervizor üsüllarından istifadə edəndə olur. Bu zaman giriş vektorlar çoxluğu və ona uyğun çıxış vektorları çoxluğu əvvəlcədən müəyyənləşdirilir. Hər bir giriş vektorunun i -ci komponenti şəbəkənin i - ci neyronuna verilən siqnala uğun gəlir. Analoji olaraq, çıxış vektorunun j - cu komponenti j - cu çıxış neyronunda alınan siqnala uyğun gəlir. X giriş vektoru və ona uyğun Y çıxış vektoru öyrətmə cütləri adlanan cüt təşkil edir. Bütün öyrətmə cütləri toplusu öyrətmə çoxluğunu yaradır. Öyrətmə prosesi zamanı neyron şəbəkənin çıxış vektorunun cari qiymətləri ilə öyrətmə çoxluğundan seçilmiş verilmiş qiymətlər arasındakı meyletmələr hesablanır. Bu qiymətləndirməyə uyğun olaraq şəbəkənin parametrlərində təshih aparılır. Neyron şəbəkələrin öyrətmə alqoritmləri arasında ən çox tətbiq edilən xətaların geriyə yayılması (*error backpropagation*) alqoritmi bu prinsiplə işləyir. Supervizor öyrətmə alqoritminə sonra daha ətraflı baxılacaqdır.

Qeyri-supervizor öyrətmə alqoritmlərində neyron şəbəkənin xarici mühitlə qarşılıqlı əlaqəsi minimuma endirilir. Bu alqoritmlərdə öyrətmə çoxluğu yalnız giriş vektorlarından ibarətdir. Burada öyrətmənin məqsədi şəbəkənin parametrlərini müvafiq qaydada seçməklə giriş vektorlar çoxluğuna xas olan qanunauyğunluqların aşkara çıxarılmasıdır.

Bu sinif alqoritmlərin ən çox yayılanı Rəqabətli öyrətmə alqoritmidir (competitive learning). Rəqabətli öyrətmənin klaster analizi ilə oxşarlığı çoxdur. Klaster analizində məsələ obrazların bir neçə klasterlər üzrə təsnif edilməsindən ibarətdir. Neyron şəbəkələrdən istifadə edərək bu məsələni belə təsvir etmək olar. Hər biri çıxış neyronlarından birinə uyğun olan giriş vektorları çoxluğu və klasterlər çoxluğu var. Şəbəkənin girişinə uyğun klasterə aid edilməli vektor daxil olan zaman onun çıxış neyronunun fəallaşmasını təmin etmək tələb olunur. Bu məsələni həll etmək üçün "hər şey qalibindir" strategiyası istifadə olunur. Belə ki, tanıma fazasında əvvəlcə verilən giriş vektorundan və çəki əmsallarından asılı olaraq təsnifedici layın neyronları öz vəziyyətlərini müəyyənləşdirmək üçün "yarış" keçirirlər. "Qalib" neyron, yəni girişlərin əmsallara vurulmuş cəminin ən böyük qiymətinə uyğun neyron "1" vəziyyətini alır, eyni zamanda başqa neyronlar "0" vəziyyətini alırlar. Eyni çıxış neyronunun şəbəkənin girişinə müxtəlif klasterlərə aid olan vektorlar daxil edilərkən fəallaşmaması üçün hər bir neyronun çəkilərinin cəminə, onu sabit kəmiyyət kimi götürməklə, məhdudiyyət qoymaq lazımdır. Öyrətmə fəal çıxış neyronları ilə fəal giriş neyronları arasındakı çəki əmsallarının artmasını və aktiv olmayan giriş düyünləri ilə əlaqələrin çəkilərinin azalmasını nəzərdə tutur.

Öyrətmə nəticəsində hər bir çıxış neyronu müəyyən klasterdən olan vektorun daxil olmasına reaksiya verəcəkdir və girişdə başqa klasterə aid olan vektor olduqda passiv olacaqdır. Lakin əvvəlcədən məhz hansı çıxış neyronunun bu klasterə uyğun olduğunu müəyyən etmək olmaz. Hər halda bu xüsusi çətinliklər törətmir. Belə ki, zərurət olduqda öyrətmədən sonra çıxış neyronlarını arzu olunan qaydada yenidən sıralamaq olar. Qeyri-supervizor öyrətməyə daha ətraflı şəkildə sonra baxılacaqdır.

Xarici mühitlə qarşılıqlı təsirin intensivliyinə görə təsdiq etmə ilə öyrətmə alqoritmləri (reinforcement learning) supervizor və qeyri-supervizor üsulları arasında aralıq vəziyyətini tuturlar.

Bu üsulun əsas prinsipi mühit (müəllim) tərəfindən təsdiq (təkzib) və ya təşviq (cəza) siqnalının (reward/penalty) olmasıdır. Əgər giriş vektoru veriləndə neyron şəbəkənin davranışı qənaətbəxş olarsa, onda ("+1") təsdiqetmə siqnalı verilir, əks təqdirdə ("0" və ya "-1") təkzibetmə siqnalı verilir. Bu zaman təsdiqetmə siqnalının yüksək tezliklə alınması məqsədi ilə şəbəkə çəki əmsallarının qiymətini dəyişir. Təsdiqetmə siqnalının tezliyi ən yüksək səviyyəyə çatmayana qədər öyrənmə davam edir. Qeyd etmək lazımdır ki, bu alqoritmlər sinfində mühit tərəfindən yalnız bir bit informasiya verilir ki, həmin informasiya şəbəkənin giriş vektorunun yaxşı və ya pis tanınmasını müəyyənləşdirir. Bu zaman şəbəkənin davranışının nə dərəcədə yaxşı və pis olması barədə informasiya olmur. Təsdiqetmə ilə öyrətmə prosedurları da bu bölmədə ətraflı izah olunacaqdır.

Neyron şəbəkələrinin öyrətmə alqoritmlərinin təsnifatındakı növbəti əlamət şəbəkənin topoloji dəyişikliklərinin xarakteri ilə əlaqədardır.

Öyrətmə prosesində seçilmiş alqoritmdən asılı olaraq həm neyronlar arasındakı əlaqələrin çəki əmsalları, həm də şəbəkənin öz quruluşu dəyişə bilər.

Öyrətmə alqoritmlərinin böyük əksəriyyəti öyrətmə gedişində şəbəkənin quruluşunun dəyişmir, yalnız əlaqələrin çəkilərini dəyişirlər. Bu alqoritmlərlə bütün öyrətmə çoxluğunda şəbəkənin qənaətbəxş davranışına nail olmaq məqsədi ilə çəki əmsallarının sazlanması aparılır. Bioloji tədqiqatlar süni neyron şəbəkələrində bu sinif prosedurların tətbiqi nəticəsində əmələ gələn proseslərlə insan beynində öyrətmə prosesində əmələ proseslərin adekvat olduğunu təsdiq edir.

Digər tərəfdən, öyrətmə prosesində neyron şəbəkənin quruluşunun dəyişməsi, bioloji cəhətdən həqiqətə tam uyğun deyil. Beyində olan belə dəyişikliklər neyronların bir hissəsinin məhv olması və onlar arasındakı əlaqələrin pozulmasına səbəb olan zədələnmə və xəstəliklər nəticəsində ola bilər. [8]-də qeyd olunur ki, insan doğulanda onun beynində tam neyron quruluşu olur (təxminən 10^9-10^{10} neyron və 10^{11}-10^{12} neyronlar arası əlaqə). Bu quruluşun dəyişilməsi yalnız degenerativ proseslərin nəticəsində olur.

Neyron şəbəkələrə dəyişən quruluşlu kimi baxan öyrətmə alqoritmləri sinfi tədqiqatlar və layihələrin nisbətən təzə və az öyrənilmiş istiqamətidir. Əlaqələrin çəkilərinin qiymətini modifikasiya edən alqoritmlərdən fərqli olaraq baxılan sinif alqoritmlər yeni elementlər və əlaqələr əlavə etməklə öyrətmə prosesində şəbəkənin quruluşunun özünü dəyişdirir. Bu sinif alqoritmlərdən ən çox tanınanı “ələ almaq ilə öyrətmə” (*recruitment learning*, “*recruit*” - rekrut sözündən) adlanan alqoritmlərdir. “Ələ alma” adı ona görə seçilmişdir ki, öyrətmə prosesində “yeni neyronu ələ alma”, yəni onun köməyi ilə tələb olunan məsələnin həll edilməsini mümkün edən şəbəkənin yaradılması üçün yeni neyronların daxil edilməsi (ələ alınması) baş verir. Ümumi şəkildə şəbəkədə iki cür neyron çoxluqları var: ələ alınmış və ya əlaqələndirilmiş və ələ alınmamış və ya sərbəst neyronlar. Əlaqələndirilmiş neyronlar elə neyronlardır ki, tanınma məsələlərinin həllində artıq iştirak edirlər və müəyyən şəkildə informasiya yükü daşıyır. Əlaqələndirilmiş neyronlar həm başqa əlaqələndirilmiş elementlərlə, həm də sərbəst neyronlarla birləşdirilə bilər. Əlaqələndirilmiş neyronlar və sərbəst elementlər arasında fərq bundan ibarətdir ki, əlaqələndirilmiş neyronların aktivlik dərəcələri toplusu şəbəkənin müəyyən məna kəsb edən vəziyyətini göstərir. Əlaqələndirilmiş elementlər sərbəst elementlərdən axırıncılar arasında əlaqələrin gücləndirilməsi nəticəsində yaranır. Ona görə də sərbəst elementlərdən yaranmış şəbəkə, “ilkin” şəbəkə adlanır [9]. Sərbəst elementlər həm əlaqələndirilmiş, həm də digər sərbəst elementlərlə əlaqəyə malik ola bilərlər.

Öyrətmə prosesində sərbəst neyronların ardıcıl olaraq “ələ alınması” baş verir, yəni cəlb edilmiş neyronlar və sərbəst neyronlardan biri (rekrut) arasında əlaqə tədricən gücləndirilir. Nəticədə sərbəst neyronlar əlaqələndirilmiş neyronlara çevrilir və onlar qoyulmuş məsələnin həllində istifadə olunur.

Girişlərə qoyulan tələbata görə öyrətmə alqoritmlərini iki sinfə ayırmaq olar. Nümunələrə görə öyrətmə kimi səciyyələnən birinci alqoritmlər sinfində uyğun öyrətmə üçün öyrətmə çoxluqları təsvir olunur. Bu zaman öyrətmə prosedurları öyrətmə çoxluğundakı nümunələrə əsaslanaraq neyron şəbəkələrin parametrlərini sazlayır.

Xətaların geriyə yayılması alqoritmi istifadə olunan halda giriş-çıxış öyrətmə vektorları cütü, rəqabətli öyrətmə halında isə yalnız giriş vektorları çoxluğu nümunə kimi çıxış edir. Nümunələr üzrə öyrətmə prosesində şəbəkə öyrətmə çoxluğunda ümumiləşdirmə aparır, yəni verilmiş nümunələrə görə həll edilən məsələnin qanunauyğunluqlarını hasil edir ki, bu da induksiyadır, başqa sözlə, xüsusidən ümumiyə doğru hərəkətdir. Bu cür ümumiləşdirmə nəticəsində şəbəkə öyrətmə çoxluğunda olmayan yeni vektorlar cütünü müəyyən dəqiqliklə tanıya bilər. Ümumiləşdirmə xassəsi öyrədilmə ilə yanaşı neyron şəbəkələrin ən əhəmiyyətli xassələrindən biridir.

Nümunələr üzrə öyrətmə alqoritmləri ümumiləşdirmənin tələb olunan səhihlik dərəcəsini təmin etmək üçün nümunələr çoxluğunda təmsiledici seçmənin verilməsini tələb edir. Bu problemi aradan qaldırmaq üçün "yeganə nümunədən istifadə etməklə öyrətmə" (*one-shot learning*) adlanan prosedur işlənilmişdir. Bu prosedur Mitçel tərəfindən izah edilmişdir [10].

Mitçel "izaha əsaslanan ümumiləşdirmə" termini daxil etmişdir. Təklif olunmuş öyrətmə proseduruna uyğun olaraq şəbəkəyə yeganə nümunə verilir və predmet sahəsi haqqında biliklərdən istifadə etməklə ümumiləşdirmə aparılır. Bu prosesin gedişində yeni "bir parça" bilik əldə edilir ki, bu öyrətmə nümunəsinin ifadə etdiyi xassə də daxil olmaqla yeni xüsusiyyət və qanunauyğunluqları təsvir edir. "İzah" termini ona görə istifadə olunmuşdur ki, öyrətmənin başlanğıcında xüsusi məqsəd konsepsiyasını təsvir etmək üçün bu nümunənin seçilməsini əsaslandıran izahlar qurulur. Bu izah hər şeydən əvvəl məqsəd konsepsiyasının müəyyənləşdirilməsində əhəmiyyətli olan nümunəyə məxsus səciyyəvi xususiyyətlərin aşkara çıxarılmasını nəzərdə tutur. Göstərilən xususiyyətlər aşkara çıxarıldıqdan sonra konsepsiyanın ümumiləşmiş tərifi qurulur. Beləliklə, konsepsiyanın tanınması yeni giriş vektorunda məqsəd konsepsiyasına məxsus xususiyyətlərin olmasının yoxlanması yolu ilə icra edilir. Şübhəsiz ki, bu alqoritmlər sinifində tədqiqat sahəsi haqqında müəyyən biliklərin olması tələb olunur. Nümunəyə körə öyrətmə üsullarında buna ehtiyac yoxdur.

6.1 bölməsində determinik və stoxastik neyron şəbəkələri izah edilmişdir [11]. Determinik neyron şəbəkələrində neyronlar determinik fəallaşma qaydalarına uyğun fəallaşır. Qeyd edildiyi kimi determinik şəbəkələrdə 4 cür fəallaşma qaydalarını bir-birindən ayırırlar: xətti, sərhəd, siqmoid (yarımxətti) və hiperbolik tangens. Neyronun fəallığı

$$y_j = \mathrm{A}\left(\sum_i w_{ji} x_i\right) \tag{6.12}$$

münasibəti ilə müəyənləşdirilir, burada y_j j-cu neyronun vəziyyəti; w_{ji} i-ci və j-cu-neyronlar arasındıkı əlaqənin çəki əmsalı; $A(\,)$ - determinik fəallaşma qaydasıdır. Göründüyü kimi, (6.12) ifadəsində bütün kəmiyyətlər determinikdir. Beləliklə, çəkilərini giriş siqnallarının məlum qiymətlərində neyronun fəallıq səviyəsini dəqiq hesablamaq olar. Bu halda öyrətmə şəbəkədəki əlaqələrin çəkisini modifikasiya edir və bununla konkret giriş vektoru verildikdə neyronların vəziyyətini yeniləşmiş çəkilərlə dəqiq təyin etmək olar.

Stoxastik neyron şəbəkələrində məsələ bir qədər başqa cürdür (məsələn, Bolsman maşınlarında). Fəallaşma qaydası burada özünü ehtimalların paylanma sıxlığı funksiyası kimi göstərir. Bu qayda vasitəsilə neyronun fəallaşma dərəcəsinin qiymətinin özü yox, neyronun "1" vəziyyətini alması ehtimalı təyin edilir

$$P_j = \frac{1}{1+e^{-\Delta E_j/\mathrm{T}}}, \text{ burada } \Delta E_j = \sum_i w_{ji} x_i\,. \tag{6.13}$$

Bu halda öyrətmə çoxluğu özünü neyronların fəallıq ehtimallarının məqsəd paylanması kimi göstərir. Öyrətmənin məqsədi imkan dairəsində elementlərin fəallaşması ehtimallarının cari paylanmasının arzu olunan paylanmaya daha çox yaxınlaşdırmaqdır. (6.13) münasibətindən körünür ki, əlaqələrin w_{ji} çəkilərini manipuliyasiya etməklə paylanmanı dəyişdirmək olar. Beləliklə, Bolsman maşınının öyrətmə prosesi çəkilərin qiymətlərini modifikasiya etmək yolu ilə giriş vektorları çoxluğu üzrə yol verilə bilən xətaların səviyyəsində neyronların fəallaşma ehtimallarının arzu olunan paylanmasının əldə edilməsindən ibarətdir.

Bolsman maşınını öyrətmə həddən çox vaxt tələb edir ki, bu da onun tətbiq sahəsini daraldır.

Bolsman maşının işləmə sürətinin artırılması üçün onun determinik analoqu təklif edilmişdir. Determinik Bolsman maşını stoxastik maşınların işinin aproksimasiya edilməsi üçün nəzərdə tutulmuşdur. Lakin bu maşın hələlik az tədqiq edilmişdir və onun geniş tətbiq olunması məsələsi açıq qalır.

Supervizor öyrətmə alqoritmləri.

Neyron şəbəkələri öyrətmə prosesi onların elə sazlanmasıdır ki, şəbəkələrin davranışı arzu olunan davranışa maksimal yaxınlaşmış olsun. Şəbəkənin davranışı dedikdə giriş siqnallarının verilən çoxluğunda neyronların müəyyən vəziyyətlərdə olması başa düşülür. Şəbəkənin davranışı onun parametrlərindən: çəki əmsallarından, neyronların fəallaşma hüdudundan və quruluşundan asılıdır. Bundan sonra fərz edəcəyik ki, öyrətmə gedişində şəbəkənin quruluşu dəyişmir. Beləliklə, çəkiləri və hüdudları modifikasiya etməklə şəbəkənin davranışını sazlamaq olar. Supervizor alqoritmlərində şəbəkənin arzu olunan davranışının öyrətmə gedişində bilavasitə verilməsindən istifadə olunur. Supervizor alqoritmlərində şəbəkənin arzu olunan davranışının təyin edilməsi prosedurlarına daha ətraflı baxaq.

Tutaq ki, n - ölçülü $X_1^n = \{x_{11}, x_{12}, ..., x_{1n}\}$ giriş vektoru verilib və onu m - ölçülü $Y_1^m = \{y_{11}, y_{12}, ..., y_{1m}\}$ çıxış vektoruna inikas etdirmək tələb olunur. $X_1^n \to Y_1^m$ inikasını təmin etmək üçün i -ci giriş neyronuna X_1^n vektorunun i -ci komponenti və hər bir j -cu çıxış neyronuna çıxış vektorunun müvafiq j -cu komponenti uyğun gələn neyron şəbəkəsi qurulur. Bu halda öyrətmənin məqsədi ondan ibarətdir ki, şəbəkənin girişinə X_1^n vektoru veriləndə imkan dairəsində daha dəqiq Y_1^m vektoru alınsın. X_1^n və Y_1^m vektorları öyrətmə cütləri adlanırlar. Ümumi halda şəbəkədə eyni çəki əmsalları və hüdudlardan istifadə etməklə $X_i^n \to Y_i^m$, $\left(i = \overline{1, N}\right)$ inikaslar çoxluğunun təmin olunması tələb olunur, burada N öyrətmə cütlərinin sayıdır. Öyrətmə cütlərinin $\left\{X^n, Y^m\right\}$ çoxluğu öyrətmə çoxluğu adlanır. Öyrətmə prosesində girişə növbə ilə X_i^n giriş vektorları daxil olurlar və çıxışda Y_i^m çıxış vektorları alınır. Öyrətmə çoxluğundan götürülən Y_i^m çıxış vektoru ilə neyron şəbəkənin çıxışında alınan uyğun vektor arasındakı fərq öyrətmə xətası adlanır. Bu xətaya uyğun olaraq şəbəkənin parametrlərinin artımlarının həm işarələri və həm də qiymətləri elə təyin olunurlar ki, öyrətmə xətasını kiçilməsinə doğru irəliləyiş təmin edilsin. Öyrətmə xətasının mümkün səviyyəsi əldə olunmayanadək öyrətmə prosesi davam etdirilir.

Supervizor öyrətmə alqoritmləri içərisində hər şeydən əvvəl xətaların geriyə yayılması (*error backpropagation*) alqoritmini xüsusi qeyd etmək lazımdır. Həmin alqoritmə bu bölmədə ətraflı baxılacaqdır. Hal-hazırda dünyada mövcud olan tətbiqi neyron şəbəkələrinin xeyli hissəsi bu alqoritmin köməyi ilə işlənilmişdir.

Bununla belə, mütəxəssislər neyron şəbəkələrin supervizor öyrətmə üsullarının bioloji həqiqətə bənzərliyi barədə olduqca pessimistdirlər. Öyrətmənin aparılması üçün məqsəd göstərişləri genererasiya edən, neyron şəbəkənin cari və arzu olunan davranışını tutuşduran və uyğun düzəlişlər edən mexanizmin insan beynində də olması olduqca az ehtimallıdır.

Buna baxmayaraq neyron şəbəkələrinə əsaslanan müxtəlif süni intellekt sistemlərinin yaradılması üçün supervizor alqoritmlərinin səmərəli olması onların geniş yayılmasını müəyyənləşdirmişdir.

Supervizor öyrətmə alqoritmlərinin bəzilərinə daha ətraflı baxaq.

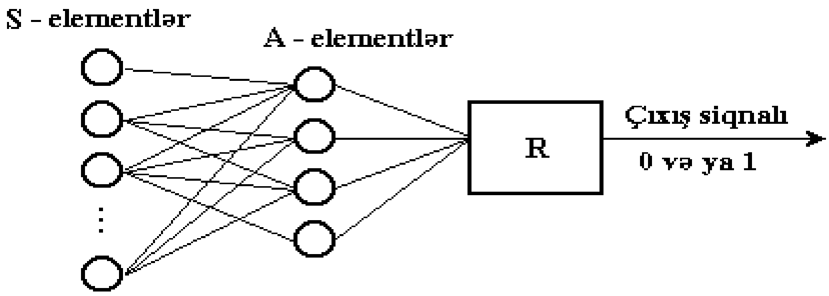
Perseptron və onu öyrətmə.

Neyron şəbəkələri arasında perseptron növlü şəbəkələr xüsusi sinif təşkil edir. Belə şəbəkələrdə laylarda yerləşən neyronlar arasında əks əlaqə yoxdur. Perseptron binar neyronlardan ibarətdir, topologiyası sadədir və adətən üç əsas komponenti birləşdirir:

1. Sensor neyronlar (torlu qişa). Bunlara giriş (tanınan) obrazlar və ya vəziyyətlər daxil olurlar. Sadə perseptronlarda (şəkil 6.15) onlar reseptor sahəsinə uyğundur və S - elementlər adlanır.

2. Torlu qişanın altçoxluğu ilə təsbit edilmiş əlaqəli neyronlar (əlamətlər detektoru) və ya A -elementləri (assosiativ elementlər) (şəkil 6.15). Bunlar bəzi S - elementlərdən siqnalı qəbul edir.

3. Əlamətlər detektoru ilə dəyişən əlaqəli binar neyronlar (həlledici elementlər). Şəkil 6.15-də onlar A - elementlərinin təsiri altında çıxış siqnalları formalaşdıran R - elementlər kimi göstərilmişdir.



Şəkil 6.15 Sadə perseptron

Həlledici elementlərin sayını perseptronun reseptorlar sahəsinə daxil olan obrazları ayırmağa lazım olan siniflərin sayına bərabər seçilirlər.

Göründüyü kimi, perseptronun modeli yalnız düz əlaqələrin olması ilə səciyyələnir. Fərz edək ki, perseptronun sensor neyronlarına hər biri həyacanlandırılmış reseptorlar (S-elementlər) toplusunu müəyyənləşdirən vəziyyətlər ardıcıllığı daxil olur. Əgər R-element yaranmış vəziyyətə “yaxşı” reaksiya verirsə, onda təşviq siqnalı verilir. Onda perseptronda “faydalı” əlaqələr güclənəcək və sensor neyronları tərəfindən qəbul edilən vəziyyətlərə perseptronun göstərdiyi səhih reaksiyanın tezliyi artacaqdır. Deməli, öyrətmə prosesinin əvvəlində əlaqələrin intensivliyi necə olursa olsun perseptron təşviq sistemi tərəfindən vadar edilən müəyyən davranışa öyrədilə bilər.

İndiyə kimi perseptronun müxtəlif öyrətmə alqoritmləri təklif olunmuşdur. Bunlardan birinə baxaq.

Perseptron onun girişinə giriş obrazlarının ardıcıl daxil edilməsi və şəbəkənin çəki əmsallarının hər bir giriş obrazı üçün arzu olunan çıxış siqnalı alınana qədər sazlanması ilə öyrədilir.

Tutaq ki, biz düzbucaqlı kartda təsvir olunmuş rəqəmlərin tək və cütlüyünü aşkara çıxarmaq üçün perseptron sistemini öyrətmək istəyirik. Kart şaquli və üfqi xətlərlə bərabər düzbucaqlı xanalara ayrılmışdır və bunların hər birinə perseptronun ayrıca girişi qarşı qoyulmuşdur. Əgər rəqəmin təsviri xananın sahəsini kəsirsə, onda bu xanaya uyğun girişə vahid, əks təqdirdə isə sıfır verilir. Sistemin bir çıxışı var və onun qiyməti girişə tək rəqəm daxil olanda 1, cüt rəqəm daxil olanda 0 olmalıdır.

Hər bir giriş obrazına elementləri 0 və 1 olan $X=(x_1,x_2,\ldots,x_n)$ vektorunu qarşı qoy-maq olar. y çıxışı X vektorunun komponentlərinin əmsallaşdırılmış cəminin hər hansı θ hüdud qiymətini aşması ilə təyin olunur

$$y=\begin{cases}1, & \sum \mathrm{x_i} w_i \geq \theta \\ 0, & \text{якс тягдирдя}\end{cases}$$

Öyrətmə aşağıdakı kimi aparılır. Girişə növbəti X obrazı verilir və ona uyğun Y çıxışı hesablanır. Əgər o doğrudursa (yəni verilmiş giriş üçün arzu olunan qiymətə uyğun gəlirsə), onda çəki əmsallarının qiymətləri dəyişdirilmir, əks halda isə bu səhv nəticələrə güclü təsir edən əlaqə çəkiləri səhvi aradan qaldırmaq məqsədilə modifikasiya edilirlər. Xətaların tam aradan qaldırılması üçün çəkiləri güclü dəyişməyə də aludə olmaq lazım deyil. Bu öyrədilmiş obrazların artıq «yaddan çıxarılmasına» gətirib çıxara bilər. Obrazları ardıcıl və ya təsadüfi olaraq verməklə, sistemi düzgün fəaliyyətə doğru addım-addım yaxınlaşdırmaqla öyrətməni bütün obrazlar üzrə qlobal aparmaq lazımdır.

Bizim sistem üçün öyrətmə prosesi aşağıdakı kimi gedəcəkdir. Tutaq ki, girişə "3" rəqəmini təsvir edən obraz daxil olur, sistemin çıxışı $y=1$-dir. Çıxış səhih olduğundan çəkilər modifikasiya edilmir. Girişə "4" obrazı daxil olanda çıxış $y=1$ olarsa, vahid girişlər daxil olan əlaqələrdə çəki əmsallarını kiçiltmək lazımdır, çünki onlar səhv nəticəyə gətirir. Analoji olaraq, əgər "3" rəqəmi 0 nəticəsini yaradırsa, onda vahid girişlər ilə bağlı çəkilər xətanı aradan qaldırmaq üçün böyüdülməlidir. Perseptronun təsvir edilmiş alqoritminin daha ümumiləşdirilmiş sxemini verək.

Addım 1. Növbəti X giriş obrazını qəbul etməli və onun üçün Y çıxışını hesablamalı.

Addım 2.

a) çıxış səhihdirsə, Addım 1-ə qayıtmalı;
b) çıxış səhih deyilsə və sıfırdırsa ($Y = 0$), hər bir w_i çəkisinə uyğun x_i girişinin qiymətini əlavə etməli;
c) çıxış səhih deyilsə və vahidə bərabərdirsə ($Y = 1$), hər bir w_i çəkisindən uyğun x_i girişinin qiymətini çıxmalı;

Addım 3. Addım 1-ə qayıtmalı.

Alqoritmdə Addım 2 δ və η kəmiyyətləri əlavə edilməklə ümumiləşdirilə bilər. $\delta = M - A$, yəni çıxışın arzu olunan (məqsəd) – M və həqiqi (aktual) – A qiymətləri arasındakı fərqdir. η çəkilərin dəyişməsini tənzim etmək üçün istifadə olunan öyrətmə sürətidir. Beləliklə, Addım 2 aşağıdakı kimi yazılır:

$$\Delta_i = \eta \delta x_i ,$$

$$w_i(n+1) = w_i(n) + \Delta_i ,$$

burada Δ_i x_i girişi ilə bağlı düzəliş, $w_i\,(n+1)$ modifikasiyadan sonra i-ci girişlə bağlı çəkinin qiyməti; $w_i(n)$ modifikasiyaya qədər i-ci girişlə bağlı çəkinin qiymətidir.

Alınan alqoritm delta-qayda adlanır və nəinki binar (0 və 1), həm də həqiqi giriş və çıxışlar üçün yarayır.

Qeyd etmək lazımdır ki, yuxarıda təqdim olunan alqoritmin sistemi düzgün fəaliyyətə gətirməsinə (əgər bu mümkündürsə) baxmayaraq, biz bunun üçün lazım öyrətmə müddətini qiymətləndirə bilmirik. Bundan başqa, bəzi hallarda çəkilərin adi seçmə yolu ilə tapılması bizi məqsədə tez çatdıra bilər. Bütün bunlarla yanaşı, əlavə etmək lazımdır ki, perseptronun hesablama qabiliyyəti heç də sonsuz deyil və ümumiyyətlə, heç də istənilən məsələ üçün həmişə həlli vermir. Həllin varlığı şərtinin yoxlanılması (giriş vektorlarının xətti separabelliyi şərti) xeyli mürəkkəb ola bilər.

Gizli layı olmayan şəbəkələr üçün supervizor öyrətmə alqoritmi.

Gizli layları olmayan, yəni giriş x_i neyronları y_j çıxış neyronlarına bilavasitə birləşdirilmiş ikilaylı şəbəkəyə baxaq (şəkil 6.16). Giriş neyronlarının fəallıq səviyyəsi daxil olunan

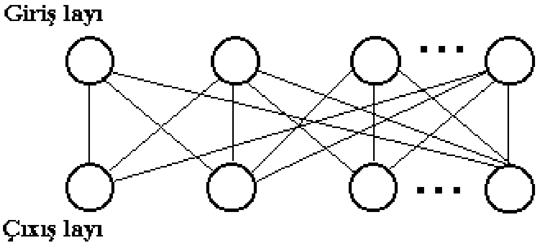
$$y_j = f\left(\sum_i w_{ji} x_i - \theta_j \right)$$

giriş vektorundan asılıdır.Burada w_{ij} i giriş neyronu ilə j çıxış neyronu arasında olan əlaqənin çəkisi, θ_j j-cu neyronun aktivlik hüdudu qiyməti; f kəsilməz hamar fəallaşma funksiyasıdır. Tutaq ki, $\{X^n, Y^m\}$ öyrətmə vektorları cütləri çoxluğu verilib. Məsələ xüsusi

X_i^n giriş vektorlarının uyğun Y_i^m çıxış vektorlarına imkan dairəsində daha səhih inikasının qurulmasından ibarətdir. Öyrətmə aparmaq üçün bizə şəbəkənin işinin korrektlik ölçüsünü təyin etmək lazımdır. Belə ölçü olaraq aşağıdakı kəmiyyət götürülə bilər:

$$E = \frac{1}{2}\sum_{j,k}(y_{jk}^* - y_{jk})^2, \ j = \overline{1,m}, \quad k = \overline{1,c}, \tag{6.14}$$

burada y_{jk}^* k-cı öyrətmə cütündən olan çıxış məqsəd vektorunun j-cu komponentinin qiyməti, y_{jk} çıxış vektorunun k-cı inikasında j-cu komponentinin həqiqi qiyməti; m çıxış vektorunun ölçüsü; c öyrətmə çoxluğunda giriş/çıxış cütlərinin sayı; E öyrətmənin ümumi xətasıdır. (6.14)-də çıxış siqnallarının həqiqi və məqsəd qiymətləri fərqinin kvadratı ona görə götürülmüşdür ki, onların işarələri müxtəlif olduqda qarşılıqlı kompensasiyadan yan keçmək mümkün olsun.



Şəkil 6.16 Gizli layı olmayan şəbəkə

Beləliklə, öyrətmə məsələsini (6.14) münasibətinin minimallaşdırılması məsələsinə gətirmək olar. Bu məsələni həll etmək üçün əvvəlcə çəki əmsallarının qiymətlərini təsadüfi olaraq vermək, xətaların alınmış qiymətinə görə istiqaməti müəyyənləşdirmək və çəkilər fəzasında xətanın minimal qiymətinə uyğun nöqtəyə tərəf irəliləmə addımını seçmək lazımdır. "Qradiyent üzrə enmə" (belə ki, xətanın qiymətinin tədricən minimuma doğru enməsi həyata keçirilir) adlanan bu hərəkət aşağıdakı münasibətlə təyin edilir

$$\Delta w_{ji} = -\varepsilon \frac{\partial E}{\partial w_{ji}} \tag{6.15}$$

Burada "öyrətmə sürəti" adlanan ε kəmiyyəti minimuma doğru hərəkətdə addımı müəyyənləşdirir. Öyrətmə sürətini çox böyük götürdükdə şəbəkə minimum nöqtəsinin "üstündən aşıb keçə" bilər. ε-nu çox kiçik götürdükdə isə öyrətmə uzun müddət davam etsə də tapılmış həll, yəni w_{ji} - lərin qiymətləri optimal qiymətlərə daha yaxın olur. $\frac{\partial E}{\partial w_{ji}}$ törəmələrini tapaq.

$$\frac{\partial E}{\partial w_{ji}} = \sum^{c} \frac{\partial E}{\partial y_i} \frac{dy_j}{dI_j} \frac{dI_j}{dw_{ji}} = \sum^{c} (y_j^* - y_j) \frac{dy_j}{dI_j} x_i \,,$$

burada $I_j = \sum_i w_{ji} x_i - \theta$ j-cu neyronun ümumi əmsallaşdırılmış girişidir. Yuxarıda verilmiş düsturda cəmləmə bütün öyrətmə cütləri üzrə aparılır.

Beləliklə,

$$\frac{\partial E}{\partial w_{ji}} = -\sum^{c} (y_j^* - y_j) \frac{\partial y_j}{\partial I_j} x_i \tag{6.16}$$

alarıq.

Onda i giriş neyronundan çıxış neyronuna əlaqələrin çəkilərinin yeniləşmiş qiymətlərini təyin edən münasibət

$$\Delta w_{ji} = \varepsilon \sum^{c} (y_j^* - y_j) \frac{\partial y_j}{\partial I_j} x_i \tag{6.17}$$

kimi yazılır və çəkilərin yeniləşməsi aşağıdakı kimi aparılır

$$w_{ji}(n+1) = w_{ji}(n) + \Delta w_{ji} \,, \tag{6.18}$$

burada $w_{ji}(n+1)$ əlaqənin öyrətmənin $(n+1)$-ci addımından sonrakı çəkisi, $w_{ji}(n)$ əlaqənin öyrətmənin n-ci addımdan sonra $(n+1)$- ci addıma qədər olan çəkisidir.

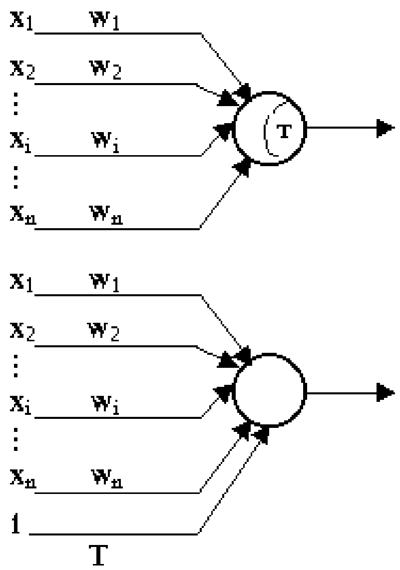
Çıxış elementlərinin hüdudlarını sazlamaq üçün, onlara sabit "1" siqnallarına qoşulmuş girişlərin çəkiləri kimi baxmaq olar (şəkil 6.17), belə ki, əvvəlcədən onların işarələrini dəyişmək lazımdır. Bunu nəzərə alaraq, biz hüdudların dəyişmə qiymətlərini təyin etmək üçün aşağıdakı münasibətləri yaza bilərik:

$$\Delta \theta_j(n+1) = \varepsilon \sum^{c} (y_j^* - y_j) \frac{\partial y_j}{\partial I_j} \tag{6.19}$$

$$\theta_j(n+1) = \theta_j(n) + \Delta \theta_j(n+1) \,, \tag{6.20}$$

burada $\theta_j(n+1)$ $(n+1)$- ci öyrətmə addımından sonrakı hüdud qiyməti, $\theta_j(n)$ $(n+1)$- ci addıma qədər hüdud qiymətidir. (6.17)-(6.20) münasibətləri gizli layları olmayan şəbəkələr üçün ən kiçik kvadratlar üsulu adlanan supervizor öyrətmə alqoritmini təyin edir [6]. $\frac{\partial y_j}{\partial I_j}$ kəmiyyəti götürülmüş fəallaşma funksiyasından asılıdır. Bu funksiya xətti olduqda $\frac{\partial y_j}{\partial I_j}$

sabit kəmiyyətdir. Fəallaşma funksiyası kimi siqmoid funksiyası götürsək, $\frac{\partial y_j}{\partial I_j} = \frac{\partial}{\partial I_j}\left(\frac{1}{1+e^{-I_j}}\right) = y_j(1-y_j)$ alarıq.



Şəkil 6.17 Hüdudların çəki əmsalları vasitəsi ilə təsviri

Ən kiçik kvadratlar üsulunun həndəsi şərhini verək. Tutaq ki n sayda giriş, m sayda çıxış neyronları var. Aydındır ki, əlaqələrin sayı və deməli, bu şəbəkədəki çəki əmsallarının sayı nm-ə bərabər olacaq. Əgər bu kəmiyyətə hüdudların sayını da əlavə etsək, onda $nm+m=(n+1)m$ alarıq. Beləliklə, ikilaylı $n\times m$ şəbəkəsində qərar qəbul etmək üçün lazım olan dəyişənlərin sayı $(n+1)m$ - ə bərabərdir. ($(n+1)m+1$) - ölçülü fəzada hər bir dəyişən üçün ox keçirək. E ümumi xətasını ölçmək üçün xüsusi ox keçirək. Çəki əmsallarının və hüdudların hər bir kombinasiyasına xətanın "hündürlüyünü" (kəmiyyətini) təyin edən hər hansı bir nöqtə uyğun gəlir. Belə nöqtələrin həndəsi çoxluğu ($(n+1)m+1$) - tərtibli hipersəth əmələ gətirir və "xəta səthi" adlanır Ən kiçik kvadratlar üsulu xəta səthi üzərində yaxınlıqda olan "çalanın" "dibi" istiqamətində daha dik enmə aparır. Enmənin başlanğıc nöqtəsi çəki əmsalları və hüdudların təsadüfi götürülmüş qiymətləri ilə təyin olunur. Xəta səthinin özü isə fəallaşma funksiyasının şəklindən asılı olaraq müxtəlif formada ola bilər. Əgər çıxış neyronlarının fəallaşma funksiyası xəttidirsə, onda bu səth piyaləyə bənzər formada olur. Bu çox- ölçülü piyaləninin dibində öyrətmə məsələsinin optimal həlli yerləşir. Qeyri-xətti funksiyalardan istifadə etdikdə xəta səthi üzərində lokal "çalalar" (yəni minimumlar) çoxluğu yarana bilər. Əgər çıxış elementləri xəttidirsə və yaxud da onların fəallaşma funksiyaları qeyri-xətti və monotondursa və "ideal" həll mövcuddursa, ən kiçik kvadratlar proseduru optimal həllin və ya ona yaxın həllin tapılmasına zəmanət verir.

Xətaların geriyə yayılması alqoritmi.

Yuxarıdakı bölmədə gizli neyronları olmayan sadə ikilaylı şəbəkələrin öyrətmə prosedurları təsvir olunmuşdur. Bu prosedur (6.14) xətasının minimal qiymətlərinin axtarışını təmin edir. Amma, təəssüf ki, kifayət qədər hesablama gücünün olmaması ucbatından ikilaylı şəbəkələr bəzi məsələləri həll edə bilmir. Belə məsələyə nümunə olaraq, hamıya yaxşı məlum olan istisnaedici "və ya" problemini göstərmək olar. Bu problemi köməkçi laylar qoşmadan sadə şəbəkə ilə həll etmək olmaz.

Çoxlaylı neyron şəbəkələrin öyrədilməsində gizli neyronlara daxil olan əlaqələrin çəki əmsallarının sazlanmasını təmin etmək üçün ən kiçik kvadratlar üsulunu ümumiləşdirmək tələb olunur. Xətaların geriyə yayılması alqoritmi [12,13] çoxlaylı şəbəkələr üçün ən kiçik kvadratlar üsulunun ümumiləşməsidir. Bu cür ümumiləşməni quran zaman gizli layların neyronları üçün xətaların ölçülərini təyin etmək problemi meydana çıxır. Bu problem xətaların ölçülərinin növbəti (sonrakı) layın xətaları vasitəsilə qiymətləndirilməsi yolu ilə həll edilir. Öyrətmənin hər bir addımında hər bir giriş/çıxış öyrətmə cütləri üçün əvvəlcə "düzünə gediş" ("*forward pass*") edilir. Bu o deməkdir ki, neyron şəbəkəyə giriş vektoru verilir. Nəticədə isə şəbəkə üzrə giriş layından çıxışa doğru istiqamətdə fəallaşma axını keçir və bütün neyronların vəziyyətləri müəyyənləşdirilir. Şəbəkənin çıxışında çıxış vektoru generasiya olunur, bu vektor məqsəd vektoru ilə müqayisə edilir və öyrətmə xətası qiymətləndirilir. Sonra çəki əmsallarının qiymətləri dəyişdirilərək bu xəta şəbəkə boyunca giriş layına doğru geriyə yayılır. Beləliklə, öyrətmə prosesi düzünə və geriyə ("*backward pass*") keçidlərin növbələşmə prosesidir, belə ki, düzünə keçiddə şəbəkənin elementlərinin vəziyyətləri təyin olunur, geriyə keçiddə isə xətanın qiymətləri yayılır və çəkilərin qiymətləri yeniləşir. Xətaların geriyə yayılması alqoritmi adı məhz buradan yaranmışdır.

Artıq qeyd edildiyi kimi, layların sayının çoxaldılması şəbəkənin hesablama gücünün artırılmasına, son nəticədə isə daha mürəkkəb inikasların qurulmasına imkan verir. Göstərmək olar ki, üçlaylı şəbəkə girişlər fəzasında praktiki olaraq istənilən qabarıq oblastları ayıra bilər. Dördüncü layın əlavə edilməsi hətta qabarıq olmayan oblastların ayrılmasını mümkün edir [1]. Beləliklə, dördlaylı neyron şəbəkənin köməyi ilə praktiki olaraq istənilən inikası qurmaq olar. Lakin bəzən çoxsaylı layları olan neyron şəbəkələrin qurulması səmərəli olmur. Gizli layların olması onların optimal istifadə olunması problemini irəli sürür. Xətaların geriyə yayılması alqoritmi gizli layların neyronlarının vəziyyətlərinin daxili interpretasiyasını qurur. Qeyd etmək lazımdır ki, gizli neyronların istifadəsi yollarının avtomatik axtarışı xeyli vaxt tələb edir və deməli, ümumi öyrətmə vaxtını artırır.

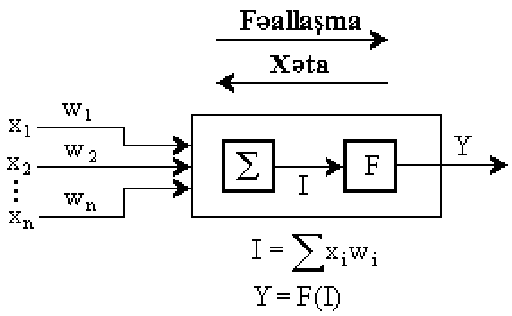
(6.15) münasibətindən göründüyü kimi w_{ji} çəkilərinin modifikasiya addımını seçmək üçün $\dfrac{\partial E}{\partial W_{ji}}$ törəməsinin qiymətini hesablamaq lazımdır. Bu qiymət öz növbəsində $\dfrac{\partial E}{\partial y_j}$ ilə təyin olunur. $\dfrac{\partial E}{\partial y_j}$ - ni qiymətləndirməkdən ötrü gizli neyronlar üçün aşağıdakı münasibət istifadə olunur

$$\frac{\partial E}{\partial y_j^{(n)}} = \sum_k \frac{\partial E}{\partial y_k^{(n+1)}} \frac{dy_k^{(n+1)}}{dI_k^{(n+1)}} \frac{dI_k^{(n+1)}}{dy_j^{(n)}} = \sum_k \frac{\partial E}{\partial y_k^{(n+1)}} \frac{dy_k^{(n+1)}}{dI_k^{(n+1)}} w_{kj} \qquad (6.21)$$

burada $y_j^{(n)}$ n-ci gizli layın j-cu neyronunun çıxışı; $y_k^{(n+1)}$ $(n+1)$-ci layın k-cı neyronunun çıxışı; $I_k^{(n+1)}$ $(n+1)$-ci laydan olan k neyronunun ümumi çəkili girişidir. Beləliklə, əgər şəbəkə m laydan ibarətdirsə, onda $\dfrac{\partial E}{\partial y_k^{(m)}}$ kəmiyyəti əvvəlcə çıxış layı üçün, bundan sonra isə ardıcıl olaraq $\dfrac{\partial E}{\partial y_k^{(m-1)}}, \dfrac{\partial E}{\partial y_k^{(m-2)}}, \ldots, \dfrac{\partial E}{\partial y_k^{(1)}}$ qiymətləri təyin edilir.

Neyron şəbəkələrin öyrətmə üsulları arasında xətaların geriyə yayılması alqoritmi ən geniş yayıldığına görə buna daha ətraflı baxaq.

Baxılan alqoritmdə istifadə olunan neyronun sxemi şəkil 6.18-də göstərilib [1].



Şəkil 6.18 Xətaların geriyə yayılması alqoritmində istifadə olunan neyron sxemi

F fəallaşma funksiyası bütün təyin oblastında diferensiallanan olmalıdır. F funksiyası olaraq adətən siqmoid funksiyasından istifadə edilir. Bu funksiyanın törəməsi, qeyd etdiyimiz kimi, aşağıdakı şəkildə olur

$$\frac{dy}{dx} = y(1-y)$$

Öyrətmə prosesindən əvvəl bütün çəki əmsallarına sıfıra yaxın təsadüfi qiymətlər vermək lazımdır. Çəkilərin başlanğıc qiymətləri bir-birinə bərabər olmamalıdır.

Şəkil 6.19-da xətaların geriyə yayılması alqoritminin blok sxemi göstərilmişdir.

Addım 1. Başlanğıc. Çəki əmsallarının ilkin qiymətləndirilməsi.

Addım 2. Öyrətmə çoxluğundan olan bütün vektorlar cütü üçün 3-6 addımlarının təkrarı. Addım 7-yə keçid.

Addım 3. Öyrətmə çoxluğundan növbəti giriş vektorunun şəbəkənin girişinə verilməsi.

Addım 4. Düzünə kediş. Laybalay şəbəkənin bütün neyronlarının vəziyyətlərinin təyini.

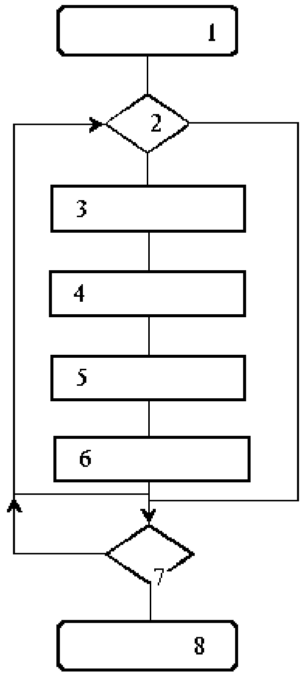
Addım 5. Neyron şəbəkənin çıxış vektoru ilə öyrətmə cütünün uyğun çıxış vektoru arasındakı meyletmənin (xətanın) hesablanması.

Addım 6. Geriyə gediş. Xətanın şəbəkə boyunca geriyə yayılması. Çəkilərin modifikasiyası.

Addım 7. Əgər ümumi öyrətmə xətası kifayət qədər kiçik deyilsə, onda Addım 2-yə qayıtma.

Addım 8. Öyrətmənin sonu.

Artıq qeyd edildiyi kimi, öyrətmənin əvvəlində çəki əmsallarına sıfır ətrafında paylanmış təsadüfi qiymətlər verilir. Öyrətmənin gedişi çəkilərin başlanğıc qiymətindən asılıdır. Uğursuz başlanğıc qiymətlərdə şəbəkə ola bilsin ki, öyrədilməsin. Bu halda öyrətmə prosesini çəki əmsallarının başqa ilkin qiymətlərində təzədən başlamaq lazımdır.



Şəkil 6.19 Xətaların geriyə yayılması alqoritminin blok sxemi

Addım 4 şəbəkənin tanıma rejimindəki işinə analojidir. Bu zaman şəbəkənin girişinə X verilir və giriş layından sonra gələn neyronlar üçün ümumi çəkili girişlər təyin edilir

$$I_j = \sum_i w_{ji} x_i ,$$

və ya vektor şəklində $I = WX$.

Bundan sonra baxılan layın neyronları $y_j = F(z_j)$ (və ya $Y = F(I)$) fəallaşma funksiyasına uyğun olaraq özlərinin çıxış siqnallarını müəyyənləşdirir. Bu layın çıxışları növbəti lay üçün girişlərdir. Neyronların vəziyyətləri ardıcıl olaraq laybalay çıxış layına çatana kimi və neyron şəbəkənin çıxışında həqiqi çıxış vektoru alınana kimi müəyyənləşdirilir.

Addım 5-də cari çıxış vektorunun həqiqi qiymətinin öyrətmə cütündən olan arzu olunan qiymətdən meyletməsi təyin edilir. Meyletmənin (xətanın) hesablanmış qiymətinə uyğun olaraq Addım 6-da əlaqələrin çəkiləri ardıcıl olaraq modifikasiya edilirlər, amma əks istiqamətdə, yəni çıxış layından girişə doğru. Giriş layına ən yaxın laya çatdıqda şəbəkənin bütün çəki əmsalları yeniləşdikdən sonra çəkilərin sazlanması prosesi dayandırılır. Şəbəkə bu yeniləşmiş parametrlər üçün ümumi öyrətmə xətasının yeni səviyyəsini müəyyən edir. Addım 7-də xəta qiymətləndirilir. Əgər bu qiymət mümkün səviyyədən yuxarıdırsa, onda öyrətmə davam etdirilir. Əgər o, qəbul edilə biləndirsə, onda proses dayandırılır və çəki əmsallarının qiymətləri tanınma (hesablama) mərhələsində istifadə etmək üçün saxlanılır.

Arzu olunan çıxış vektoru öyrətmə çoxluğundan birbaşa verildiyindən hər bir çıxış neyronu üçün xətanın təyin edilməsi

$$e_j = y_j^* - y_j \tag{6.22}$$

düsturu ilə yerinə yetirilir, burada e_j j -cu çıxış neyronunun xətası; y_j^* arzuolunan çıxış vektorunun j -cu komponentinin qiyməti; y_j həqiqi çıxış vektorunun j -cu komponentinin qiymətidir (j -cu neyronun cari çıxışı). Bu kəmiyyət fəallaşma funksiyasının birinci törəməsinə $f'(I)$ vurulur

$$\delta_j = (y_j^* - y_j)\, f'(I_j)$$

Əgər f fəallaşma funksiyası siqmoid funksiyasıdırsa, onda ifadə aşağıdakı şəkli alar.

$$\delta_j = (y_j^* - y_j)\ (1 - y_j) \tag{6.23}$$

(6.23) qiymətləndirməsinin köməyi ilə axırıncıdan əvvəlki və axırıncı laylar arasında əlaqələrin çəki əmsallarının artım kəmiyyətlərini təyin etmək olar və ya sadəcə olaraq, çıxış neyronlarının çəkilərini təzələmək olar. Bu artım kəmiyyətləri aşağıdakı kimi təyin edilir.

$$\Delta w_{ji} = \varepsilon \delta_j x_i , \tag{6.24}$$

burada Δw_{ji} j -cu çıxış neyronunun onun i -ci girişindəki çəki əmsalının artımı; x_i əvvəlki layın i -ci neyronunun çıxışı olub j -cu çıxış neyronunun girişi; ε öyrətmə sürətidir.

ε $[0.01; 1.00]$ parçasından seçilir. Çəkilərin təzələnmiş qiymətləri aşağıdakı kimi təyin edilir

$$w_{ji}(n+1) = w_{ji}(n) + \Delta w_{ji}, \tag{6.25}$$

burada $w_{ji}(n+1)$ – öyrətmənin $(n+1)$-ci addımından sonra çəkinin qiyməti; $w_{ji}(n)$ çəki əmsalının $(n+1)$-ci addıma qədər olan qiymətidir.

Beləliklə, çoxsaylı neyron şəbəkələrinin axırıncı layının öyrədilməsi üçün (6.22)-(6.25) ifadələrini aldıq. Bu alqoritmdən gizli layların öyrədilməsində istifadə etməyə cəhd göstəriləndə gizli neyronlar üçün (6.22) kəmiyyətlərinin təyin edilməsi problemi çıxır. Doğrudan da, əgər y_j^*-nın qiymətləri əvvəlcədən öyrətmə çoxluğundan verilmişsə, gizli neyronların çıxışlarının arzu olunan qiymətləri öyrətməyə qədər məlum ola bilməz. Bu fakt uzun müddət (Rummelhart və həmkarları xətaların geriyə yayılması alqoritminin ideyasını təklif edənə qədər) çoxlaylı şəbəkələr üçün öyrətmə alqoritminin işlənilməsində əngəllər törətmişdir.

Düzünə gedişdə hər bir layın hər bir neyronu öyrətmənin gedişində özünün xəta payını qoşur. Gizli neyronların son nəticəyə gətirdiyi xətanı təyin etmək üçün çıxış layının xətasına əsaslanmaq lazımdır. Beləliklə, axırıncıdan əvvəlki layın neyronları üçün xətanın miqdarı, çıxış layının xətasından istifadə etməklə təyin edilir. Sonra, k layının bütün neyronları üçün xətalar ($k+1$)-ci layın neyronlarının xətaları vasitəsilə təyin edilə bilər. Bu qayda ilə, xətalar elə bil ki, şəbəkə boyunca əksinə yayılır.

Çoxsaylı neyron şəbəkələrinin k-cı və ($k+1$)-ci laylarına baxaq. Hesab edəcəyik ki, (k+1)-ci layın neyronlarının xətası məlumdur. Onda bu layın bütün neyronları üçün $\delta_j^{(k+1)}$ kəmiyyətini təyin etmək olar

$$\delta_j^{(k+1)} = e_j^{(k+1)} y_j^{(k+1)} (1 - y_j^{(k+1)}) \tag{6.26}$$

burada $e_j^{(k+1)}$ ($k+1$)-ci layda j-cu neyronun xətası; $y_j^{(k+1)}$ ($k+1$)-ci layın j-cu neyronunun vəziyyətidir.

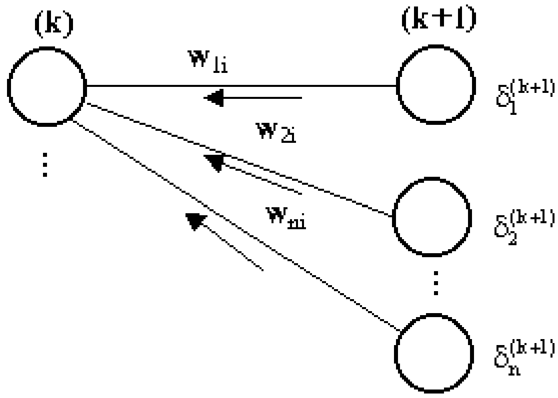
k-cı layın i-ci neyronunun xətasını təyin etmək üçün (6.26) qiymətlərini əmsallaşdırılmış əlaqələr üzrə ($k+1$)-ci layın neyronlarından i-ci neyrona kimi yazaq (şəkil 6.20). Bununla, biz k-cı layda i neyronu üçün xətanın qiymətini alarıq

$$e_i^{(k)} = \sum_j \delta_j^{(k+1)} w_{ji}^{(k+1)} = \sum_j e_j^{(k+1)} y_j^{(k+1)} (1 - y_j^{(k+1)}) w_{ji}^{(k+1)}$$

$\delta_i^{(k)}$ kəmiyyətləri

$$\delta_i^{(k)} = y_i^{(k)} (1 - y_i^{(k)}) \sum_{\mathrm{j}} \delta_j^{(k+1)} w_{ji}^{(k+1)} \tag{6.27}$$

kimi təyin olunur, burada $y_i^{(k)}$ k-cı layın i-ci neyronunun çıxışı; $w_{ji}^{(k+1)}$ k-cı layın i-ci neyronunun girişində ($k+1$)-ci laydan olan j-cu neyronun çəkisidir.



Şəkil 6.20 Xətaların geriyə yayılması

Sonra, çıxış layının neyronları üçün olduğu kimi, k layının neyronlarının çəkilərinin artım qiymətləri təyin olunur

$$\Delta w_{il}^{(k)} = \varepsilon \delta_i^{(k)} x_l^{(k)} \tag{6.28}$$

və nəhayət

$$w_{il}^{(k)}(n+1) = w_{il}^{(k)}(n) + \Delta w_{il}^{(k)}, \tag{6.29}$$

burada $x_l^{(k)}$ i-ci neyronun l-ci girişi; $w_{il}^{(k)}(n+1)$ k layında i neyronunun l girişindəki çəkisi (bu giriş $(n+1)$-ci öyrətmə addımında $(k\text{-}l)$-ci layın l neyronunun çıxışıdır); $w_{il}^{(k)}(n)$ həmin çəkinin n-ci addımdan sonrakı qiymətidir.

Bir çox məsələlərin həllində y neyronunun $I_j = \sum_i w_{ji} x_i + \theta_j$ ümumi əmsallaşdırılmış giriş düsturunda hüdud kəmiyyətindən istifadə edilməsi çox faydalıdır. Hüdud kəmiyyətindən istifadə etmək öyrətmə sürətini xeyli artıra bilər. Yerdəyişmənin öyrədilməsini təmin etmək üçün yuxarıda qeyd etdiyimiz kimi onu həmişə “1” siqnalı olan girişin çəkisi kimi təsəvvür etmək olar. Onda k layında i neyronunun $\theta_i^{(k)}$ aktivlik hüdudunu hesablamaq üçün aşağıdakı düsturlardan istifadə etmək olar

$$\Delta \theta_i^{(k)} = \varepsilon \delta_i^{(k)} \tag{6.30}$$

$$\theta_i^{(k)}(n+1) = \theta_i^{(k)} + \Delta \theta_i^{(k)} \tag{6.31}$$

Əgər şəbəkə m laydan ibarətdirsə, onda çıxış layının q neyronunun xətası

$$e_q^{(m)} = y_q^{*(m)} - y_q^{(m)} \tag{6.32}$$

olacaqdır.

Beləliklə, (6.26)-(6.32) münasibətləri xətaların geriyə yayılması alqoritmini verir. Fikir verin ki, ε şəbəkənin bütün layları üçün öyrətmənin gedişində sabit qalır.

Nəzərdə saxlamaq lazımdır ki, öyrətmə prosesində növbəti öyrətmə cütü seçilir, çəkilərdə düzəliş aparılır, sonra isə o biri cüt seçilir və s. Beləliklə, hər bir öyrətmə cütü üçün hər addımda Δw_{ji} kəmiyyəti yalnız bir dəfə təyin olunur.

Vektorların bir cütü üçün bütün öyrətmə prosesini axıra kimi aparıb, bundan sonra o biri cütə keçmək düzgün olmazdı. Bu halda, başqa cütə keçəndə birinci cüt üçün alınmış çəki "korlanır". Məsələn, əgər biz neyron şəbəkə ilə əlifbanın hərflərini tanımaq üçün "A" hərfindən başlayaraq neyron şəbəkənin onu tanımağı axıra qədər öyrətsək, bundan sonra həmin qayda ilə "B" hərfini tanımağı öyrədəndə şəbəkə "A" hərfini "yaddan çıxarar", çünki "A"-ya uyğun çəkilər artıq yeniləşmiş olacaq. Xətaların geriyə yayılması alqoritminin çatışmayan cəhətlərindən biri, artıq öyrədilmiş şəbəkəyə yeni giriş-çıxış inikası əlavə etmək lazım gəldikdə bütün öyrətmə prosesinin yeni öyrətmə cütü (və ya cütləri) əlavə edilmiş ilkin öyrətmə çoxluğu ilə əvvəldən başlamasıdır.

Öyrətmə prosesinin səmərəliliyini artırmaq üçün xətaların geriyə yayılması alqoritminin müxtəlif variantları işlənilmişdir.

Rummelhart, Xinton və Vilyams [13] xətanın geriyə yayılması alqoritminin tətbiqi zamanı dayanıqlığı artırmaq və öyrətmə müddətini azaltmaq üçün "momentlər" üsulu adlanan üsul təklif etmişlər. Onun mahiyyəti belədir. Gizli laylar və qeyri-xətti fəallaşma funksiyasından istifadə edildikdə xətalar səthinin həndəsi təsvirində qabarıqlıqlar və çökəkliklər ("təpələr" və "çalalar") olur. Öyrətmənin əvvəlində şəbəkə bu səthin nöqtələrindən birində olur. Öyrətmə prosesində yaxınlıqdakı çala tərəfə (lokal minimuma tərəf) hərəkət (enmə) baş verir. Bu zaman hərəkət addımı öyrətmənin seçilmiş sürətindən asılıdır. Əgər öyrətmə sürəti çox böyük olarsa, onda şəbəkə yaxınlıqdakı çalanı (əgər bu çala kifayət qədər dar olarsa) ötüb keçə bilər. Şərh olunan üsulda cari çəkiləri tapmaq üçün çəkilərin əvvəlki düzəliş qiymətlərindən istifadə etmək təklif olunur ki, bu da şəbəkəni bu çökəkliklərin dibinə enməyə məcbur edə bilər. Çəkilərdə düzəlişlər (korreksiya) aparmaq üçün

$$\Delta w_{ji}(n+1) = \varepsilon\delta_j x_i + \mu\Delta w_{ji}(n) \tag{6.33}$$

$$w_{ji}(n+1) = w_{ji}(n) + \Delta w_{ji}(n+1) \tag{6.34}$$

münasibətlərindən istifadə edilir, burada μ moment adlanır və 0. 9-a yaxın seçilir.

Öyrətmədə eksponensial hamarlamadan istifadə aşağıdakı münasibətlərə gətirir

$$\Delta w_{ji}(n+1) = \alpha\Delta w_{ji}(n) + (1-\alpha)\delta_j x_i \tag{6.35}$$

$$w_{ji}(n+1) = w_{ji}(n) + \varepsilon\Delta w_{ji}(n+1) \tag{6.36}$$

(6.35) düsturu Secnovski və Rozenberq [10] tərəfindən təklif olunmuşdur. Burada α hamarlama əmsalıdır. Göründüyü kimi, $\alpha = 0$ olduqda hamarlama iştirak etmir, (6.35)-(6.36) münasibətləri özünü geriyə yayılma baza alqortmi kimi göstərir. $\alpha = 1$ olduqda

$\Delta w_{ji}(n+1)=\Delta w_{ji}(n)$ olur, yəni çəkilərin əvvəlki dəyişmələri təkrar olunur. α 0 və 1 arasında olduqda α -ya mütənasib hamarlama gedir.

Öyrətmə sürətinin artırılması üçün fəallaşma funksiyasının ikinci törəməsindən istifadə etmək [15]-də təklif olunur. Fəallaşma funksiyasının ikinci törəməsindən istifadə edilən üsul xətaların geriyə yayılmasının ikinci tərtib alqoritmi adlanır. İsbat olunub ki, ikinci tərtibdən yuxarı törəmələrdən istifadə etmək öyrətmə vaxtının yaxşılaşdırılmasına gətirmir. Onu da qeyd edək ki, ikinci törəmənin hesablanması zərurəti əlavə maşın vaxtı tələb edir.

Xinton [6] xətaların geriyə paylanması alqoritminə ehtimal nəzəriyyəsi nöqteyi nəzərindən baxmışdır. Verilmiş hər bir giriş vektorun verilən çıxış vektoru üçün şərti ehtimalların paylanması qanununa tabe olur. Bu yanaşmada xətaların geriyə yayılması üsulu məqsəd (arzu olunan) paylanmalarının generasiyası ehtimalının maksimallaşdırılmasını tələb edir.

Bolsman maşınlarında olduğu kimi fərz edək ki, çıxış neyronları “0” və “1” kimi iki vəziyyət ala bilər və ikilik çıxış vektoruna həqiqi vektor uyğundur, belə ki, bu vektor özünü çıxış elementlərinin işədüşmə ehtimalları kimi göstərir. Onda göstərmək olar ki, verilmiş öyrətmə çoxluğunda məqsəd vektorların dəqiq əldə edilməsi ehtimalı o vaxt maksimal olar ki, neyronların məqsəd və həqiqi vəziyyətlərinin ehtimal paylanmalarının

$$C=-\sum_{j,k} p^*_{j,k}\log_2(p_{j,k})+(1-p^*_{j,k})\log_2(1-p_{j,k}),$$

şərti entropiyası minimal olsun, burada $p^*_{j,k}$ k halında y neyronu üçün arzu olunan ehtimaldır, $p_{j,k}$ - isə həqiqi ehtimaldır.

Beləliklə, bu şərhdə öyrətmə xətası kimi şərti entropiyanın qiymətləndirilməsindən istifadə olunur. Bu, praktikada çıxış elementlər sıfıra çox yaxın siqnallar verdikdə (lakin həqiqətdə onların məqsəd vəziyyətləri vahidə bərabərdir) yaranan çətinlikləri aradan qaldırmağa imkan verir.

Xətaların geriyə yayılması alqoritminin imkanlarını nümayiş etdirmək üçün Xinton [6] 24 müxtəlif adam arasında qohumluq əlaqələrinin neyron şəbəkələri ilə öyrənilməsi üzrə təcrübə qoymuşdur. Bu adamlar 2 soyad ağaclarına aiddirlər. Bu ağaclardan biri ingilis ailəsini, digəri isə italyan ailəsini təmsil edir. Soyad ağacı haqqında informasiya

(<adam 1>, <münasibət>, <adam 2>)

üçlüyü şəklində təmsil edilmişdir. Öyrətmə üçün 5 laylı şəbəkədən istifadə edilmişdir. Bu şəbəkənin girişləri iki yerə bölünür. Bir girişə “adam 1”, o birinə “münasibət” verilir. Şəbəkənin girişlərinə üçlüyün birinci 2 elementini verdikdə şəbəkənin çıxışında 3-cü element göstərilməlidir.

Öyrətmə üçün mümkün 104 müxtəlif münasibətdən 100-ü istifadə edilmişdir. Öyrətmədən sonra şəbəkə 100 münasibətin hamısını, hətta öyrətmə çoxluğuna daxil edilməyən 4 əlavə münasibəti səhih olaraq tanımışdır ki, bu da neyron şəbəkələrə xas olan ümumiləşdirmə xüsusiyyətini nümayiş etdirir.

Xətaların geriyə yayılması alqoritminin əsas çatışmayan cəhəti şəbəkənin lokal minimuma düşməsinin mümkün olmasıdır, yəni elə lokal həll alınır ki, bu həll qənaətbəxş olmaya bilər. Həndəsi baxımdan bu onu göstərir ki, xəta səthi üzərində yaxınlıqdakı çökəklik o qədər də dərin deyil, lakin şəbəkə bura düşərək ondan çıxa bilmir.

Bu problemin həlli yollarından biri tələb olunandan xeyli çox sayda gizli neyronlardan istifadə etməkdir. Bu lokal minimumlardan xilas olmağa kömək edir.

Baxılan öyrətmə alqoritminin digər çatışmayan cəhəti onun sürətinin nisbətən az olmasıdır. N neyronu olan şəbəkəni M inikasla öyrətmək istəsək və bu zaman L öyrətmə addımından istifadə etsək, onda ümumi öyrətmə vaxtı $M \cdot N \cdot L$ tərtibli olacaq. İnikasların sayının neyronların və addımların sayına təxminən bərabər olduğunu fərz etsək, onda çox kobud qiymətləndirmə ilə öyrətmə vaxtı $\mathrm{O}(N^3)$ tərtibinə mütənasibdir. Paralel hesablamalardan istifadə etdikdə, yəni bütün neyronlar paralel işləyərsə bu qiymətləndirmə $\mathrm{O}(N^2)$ -a qədər azalır [2,6].

Göstərilən çatışmazlıqlarına baxmayaraq xətanın geriyə yayılması alqoritmi tətbiqi neyrokompütiq sistemlərində geniş istifadə edilir.

Qeyri-supervizor öyrətmə alqortmi. Qeyri-supervizor öyrətmə alqoritmlərində öyrətmə çoxluğu yalnız giriş vektorlarından ibarətdir. Bu üsulların əsas xüsusiyyəti xətanı təyin etməyin mürəkkəb olmasıdır, çünki arzu olunan çıxış vektorları bir başa göstərilmir. Bioloji nöqteyi nəzərdən supervizor alqoritmlərinə, məsələn, xətanın geriyə yayılması alqoritminə nisbətən qeyri-supervizor alqoritmləri daha həqiqətə bənzərdir. Bu bölmədə Hebb öyrətmə alqoritminə (Nebbian learning) və Rəqabətli öyrətmə alqoritminə baxacağıq.

Hebb alqoritmi neyron şəbəkələrin öyrətmə alqoritmlərindən birincisidir. Bu alqoritm eyni zamanda aktivləşən neyronlar arasında əlaqələrin gücləndirilməsi ideyasına əsaslanmışdır. Hebbin baza alqoritmində bir sıra çatışmayan cəhətlər var. Bunlar bu bölmədə təsvir olunmuşdur.

Rəqabətli öyrətmə qeyri-supervizor alqoritmləri sinfində ən məşhur üsuldur. Bu üsulun klaster analizi ilə ümumilikləri çoxdur. Burada çıxış neyronlarına klasterlər uyğun qoyulur və şəbəkəyə girişi vektoru verildikdə, bu neyronlardan hər biri bu vektorda uyğun sinfə xas olan xüsusiyyətləri üzə çıxarmağa çalışır. Bununla elə bil ki, "yarış" keçirilir. Uyğun sinfi xarakterizə edən əlamətləri ən yaxşı meydana çıxaran neyron qalib olur və aktivləşir, bununla belə, "məğlubiyyətə" uğramış qalan neyronlar "0" vəziyyətinə keçir. Bununla da hər şey qalibindir (*winner-takes-all*) adlanan strategiya reallaşdırılır.

Hebb öyrətmə alqoritmi. Hebb fiziologiyanın və psixologiyanın nailiyyətlərinə əsaslanaraq neyron şəbəkələrin öyrədilməsi üçün sadə ideya təklif etmişdir. Bunun məğzi ondan ibarətdir ki, iki neyronun eyni zamanda fəallaşması onlar arasındakı əlaqənin güclənməsinə səbəb olur. Biologiyadakı bir çox tədqiqatlar [17] Hebbin ideyasını təsdiq edir. Bu ideyaya əsaslanan öyrətmə alqoritmi neyron şəbəkələrin öyrətmə üsullarının işlənməsində başlanğıc nöqtədir.

Alqoritmin məğzini təsvir etmək üçün neyron-mənbələr və neyron-qəbuledicilərdən ibarət olan şəbəkəyə baxaq. Neyron-mənbələr bilavasitə qəbuledicilərə birləşdirilmişdir. Neyron-mənbələrin fəallaşma səviyyələrindən təşkil edilmiş X vektoruna və neyron-qəbuledicilərin çıxış siqnallarından yaradılmış Y vektoruna baxaq. Hər bir neyron-mənbə hər bir qəbuledici neyrona çəki əmsalı w_{ji} olan əlaqəylə birləşdirilmişdir. w_{ji} kəmiyyətləri toplusu W matrisini təşkil edir. Deməli, bu halda öyrətmə prosesi W matrisinin elementlərinin təshih edilməsinə gətirilir. Matrisin hər bir elementinin artımı

$$\Delta w_{ij} = \varepsilon x_i y_j \tag{6.37}$$

düsturu ilə təyin olunur, burada ε öyrətmə sürəti olub sabit kəmiyyətdir. (6.37) ifadəsinin vektor yazılışı $\Delta W = \varepsilon YX^T$ kimi olar, burada X^T transponirə olunmuş X matrisidir (belə ki, istənilən vektora matris kimi baxmaq olar); ΔW çəki əmsallarının Δw_{ij} artımlarından təşkil edilmiş matrisdir.

$(n+1)$ -ci öyrətmə addımında çəki əmsallarının yeniləşməsi üçün

$$w_{ji}(n+1) = w_{ji}(n) + \Delta w_{ji}(n+1) \tag{6.38}$$

və ya

$$W(n+1) = W(n) + \Delta W(n+1)$$

münasibətlərindən istifadə olunur. (6.37), (6.38) münasibətləri Hebb öyrətmə alqoritmini verir.

Bir $X_1 \to Y_1$ inikasının təmin olunması və öyrətmə sürəti $\varepsilon = 1$ olan hala baxaq. Şəbəkəni bu inikasa öyrədəndə W matrisinin bütün elementlərinin başlanğıc qiymətlərinin 0-a bərabər olduğunu fərz edərək, onların yeniləşmiş qiymətləri

$$\Delta W = Y_1 X_1^T\ , \tag{6.39}$$

$$W = W^0 + \Delta W \tag{6.40}$$

düsturları ilə hesablanır, burada W^0 öyrətməyə qədər olan W matrisidir. (6.39), (6.40) düsturlarından görünür ki,

$$W = YX^T$$

Şəbəkənin girişinə X_1 vektoru verildikdə, çıxışda ona uyğun Y_1 vektoru alınır. Lakin ümumi halda, W matrisində $X \to Y$ assosiasiyalar çoxluğunun yadda saxlanılması tələb olunur. Bunun üçün Hebb alqoritmi bütün assosiasiyalar üçün iterativ olaraq tətbiq olunur. Bu zaman W matrisinin hər bir elementi bütün assosiasiyaların yerinə yetirilməsinin təmin edilməsində iştirak edir, deməli, W matrisi özünü assosiativ paylanmış yaddaş şəklində göstərir.

Bununla belə, Hebb alqoritminin ciddi bir qüsuru var. Bu qüsur ondan ibarətdir ki, assosiasiyanın səhih yerinə yetirilməsini təmin etmək üçün bütün vektorların ortoqonal olması tələb olunur, yəni

$$X_1 X_2 X_3 \cdots X_n = 0 \tag{6.41}$$

münasibəti ödənilməlidir, burada sol tərəfdə bütün giriş vektorlarının skalyar hasili göstərilmişdir. Əks halda şəbəkənin işində (6.41) məhdudiyyətinin yerinə yetirilməməsindən irəli gələn xəta müşahidə olunacaq. Əgər giriş vektorları $X_1, X_2, X_3, \cdots, X_n$ ortoqonal deyillərsə və onlara $Y_1, Y_2, Y_3, \cdots, Y_n$ vektorları uyğundursa, onda şəbəkənin girişinə hər hansı bir X_i giriş vektoru verildikdə çıxışda $Y_1, Y_2, Y_3, \cdots, Y_n$ vektorlarının xətt kombinasiyası olan Y vektoru meydana çıxacaqdır, yəni

$$Y = \alpha_1 Y_1 + \alpha_2 Y_2 + ... + \alpha_n Y_n = \sum_j \alpha_j Y_j$$

M -ölçülü fəzada yalnız M sayda ortoqonal vektor olduğunu nəzərə alsaq, onda hökm etmək olar ki, şəbəkə bu halda M -dən çox olmayan assosiasiyaları gerçəkləşdirə bilər.

Hebb öyrətməsinin çatışmamazlıqlarına baxmayaraq, bu alqoritm neyron şəbəkələrin öyrətmə üsullarının inkişafında mühüm rol oynamışdır.

Hebb öyrətmə alqoritmi ona görə qeyri-supervizor üsulları sinfinə aid edilmişdir ki, onda öyrətmə gedişində yalnız iki neyronun qarşılıqlı əlaqəsi haqqında olan informasiyadan istifadə edilir, supervizor və ya "müəllim" tərəfindən müdaxilə olmur.

Rəqabətli öyrətmə alqoritmi.

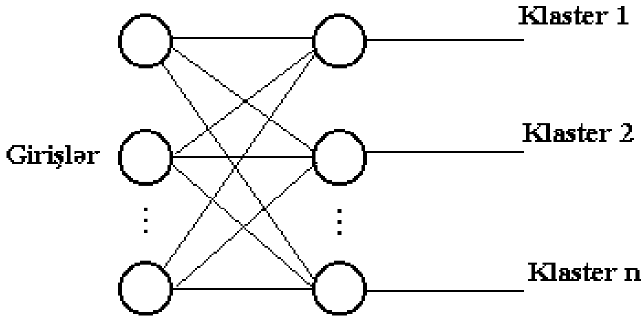
Artıq qeyd edildiyi kimi qeyri-supervizor alqoritmlərində xarici qlobal supervizor ilə qarşılıqlı əlaqə yoxdur. Öyrətmə çoxluğu yalnız giriş vektorları çoxluğundan ibarətdir, şəbəkənin arzu olunan çıxış vektorları verilmir. Qeyri-supervizor üsullarından ən məşhur və geniş yayılmış olanı rəqabətli öyrətmədə məqsəd daxil olan giriş vektorlarını bir neçə klaster üzrə təsnif etməkdir (siniflərə ayırmaqdır). Beləliklə, rəqabətli öyrətmə klaster analizi ilə sıx əlaqədardır. Klaster analizində N klasterə baxılır və daxil olan vektorun (obrazın) mövcud klasterlərdən birinə aid edilməsi tələb olunur.

Rəqabətli öyrətmədən istifadə edən neyron şəbəkə ona daxil olan bir-birinə kifayət qədər yaxın giriş vektorlarını bir klasterə aid edir. Əks halda vektorlar müxtəlif klasterlərə aid edilir.

Rəqabətli öyrətmə üsullarından biri Rummelhart və Zipser [15] tərəfindən tədqiq edilmişdir. Onlar öz işlərində səciyyəvi xüsusiyyətləri tanımaq üçün nəzərdə tutulmuş, bir və ya bir neçə layı olan əks əlaqəsiz şəbəkələrə baxmışlar.

Tanıma laylarında yerləşən neyronlar daxil olan giriş vektorunda onlara uyğun klasterlərə məxsus səciyyəvi əlamətləri aşkara çıxarmağa çalışır. Bir klasteri digərindən ayırd etmək üçün çəki əmsalları kifayət qədər müxtəlif olduğundan neyronların uğur qazanmaya "cəhd"ləri də müxtəlif olur. Haqlı olaraq "yarış" adlandırılan bu prosesin gedişində ən çox uğur qazanmış neyronlar qalib sayılır. Klasteri aktual giriş vektoruna ən yaxın olan neyron qalib neyrondur. Yarışdan sonra qalib neyron daxil olan vektoru tanımaq iddiasını başqa neyronlara qadağan edir, yəni yalnız onun bu vektoru tanımaq hüququ var. Təsvir olunan proses "hər şey qalibindir" strategiyasını həyata keçirir (winner takes all).

Tanıyan neyronun yalnız ona uyğun olan vektor girişə daxil fəallaşması üçün mühdudiyyət qoyulmalıdır. Baxılan neyronu əvvəlki layın neyronları ilə birləşdirən əlaqələrin çəkiləri cəminin sabit bir kəmiyyət olması tələbini məhdudiyyət kimi qəbul etmək olar. Şərh olunan işdə məhdudiyyət kimi neyronun çəkilərinin cəminin vahidə bərabər olması qəbul edilmişdir. Sadəlik xatirinə şəkil 6.21-də təsvir edilmiş ikilaylı şəbəkəyə baxaq. Hər bir çıxış vektoru hər hansı bir klasterlə əlaqəlidir. Şəbəkənin girişinə X vektoru veriləndə çıxış (tanıyan) neyronlarının yarışa qədər fəallaşması $y_j = \sum_i w_{ji} x_i$ ümumi əmsallaşdırılmış giriş düsturu ilə hesablanır. Burada, əvvəldə olduğu kimi, x_i i-ci giriş neyronu; y_j j-cu klasterə uyğun j-cu neyron; w_{ji} i-ci giriş neyronu ilə j-cu çıxış neyronu arasındakı əlaqənin çəkisidir.



Şəkil 6.21 Rəqabətli öyrətmə aparmaq üçün şəbəkə

Öyrətmə prosesinin başlanmasına qədər giriş vektorunu elə normallaşdırmaq lazımdır ki, onun ümumi uzunluğu vahidə bərabər olsun

$$x_i' = \frac{x_i}{\sqrt{\sum_j x_j^2}},$$

burada x_i' giriş vektorunun i-ci komponentinin normallaşmadan sonrakı qiymətidir. Rummelhart və Zipser öz işlərində X giriş vektoru binar olan hala baxmışlar. Beləliklə, yarışda qalib çıxan neyron müəyyənləşdiriləna qədər çıxış neyronlarının aktivlik səviyyələri $0 \div 1$ aralığında olur. Bundan sonra, ən yüksək fəallaşma səviyyəsinə malik neyron qalib olur, yəni $y_w = \max(y_1, y_2, ..., y_m)$. Burada y_w qalib neyrondur. Qalib neyronun fəallaşma səviyyəsi “1” olur, o biri neyronlar isə sıfır vəziyyətinə keçir, yəni

$$y_w = 1,$$
$$y_i = 0, \ i \neq w.$$

Beləliklə, yarışı yalnız bir neyron udur və burada ikinci, üçüncü və başqa yerlər anlayışı yoxdur. Bütün qalan neyronlar onların aktivlik səviyyələrinin paylanmasından asılı olmayaraq uduzmuş sayılır.

Öyrətmə prosesində girişə öyrətmə çoxluğundan vektor verilir və bundan sonra yarış başlanır. Qalib müəyyənləşdirildikdən sonra, bu neyronu aktiv giriş neyronları ilə birləşdirən əlaqələr gücləndirilir. Rummelhart və Zipser əlaqələrin çəki əmsallarını təshih etmək üçün aşağıdakı münasibətləri təklif etmişlər

$$\Delta w_{ji} = g\left(\frac{x_{ik}}{n_k} w_{ji}\right), \tag{6.42}$$

$$w_{ji}(n+1) = w_{ji}(n) + \Delta w_{ji} \tag{6.43}$$

Burada x_{ik} öyrətmə çoxluğundan olan k vektorunun i -ci komponenti; n_k n -ci vektor veriləndə aktiv giriş neyronlarının sayı; g sabitdir, öyrətmə sürətini göstərir.

(6.42), (6.43) münasibətlərindən göründüyü kimi, hər bir öyrətmə addımında qalib neyronun çəkilər vektoru girişə daxil olan vektora yaxınlaşır. Öyrətmədən sonra, girişə yeni vektor verildikdə hansı neyronun çəki əmsallarından təşkil edilmiş vektor giriş vektoruna daha yaxın olsa, o neyron aktivləşəcək. Bunu aşağıdakı mühakimələrdən başa düşmək olar.

Tutaq ki, girişə X vektoru verilir. Bütün çıxış neyronları üçün $\sum w_{ji}x_i$ qiymətinin hesablanması faktiki olaraq onu göstərir ki, çıxış neyronlarının hər biri üçün onların çəki vektorları ilə X giriş vektorunun skalyar hasili hesablanır, yəni yarışın gedişində j -cu neyronun aktivlik səviyyəsi

$$\mathrm{y_i} = \sum_{\mathrm{i}} \mathrm{w_{ji}x_i} = |\mathrm{W_j}||\mathrm{X}|\cos(\mathrm{W_j}, \mathrm{X}) \tag{6.44}$$

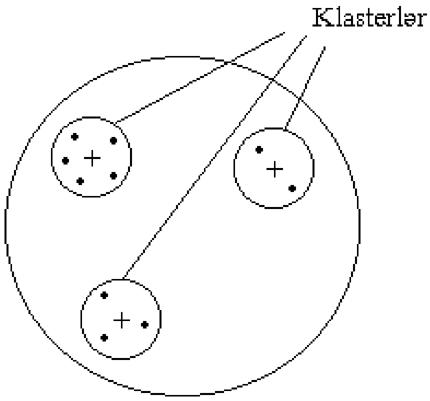
olacaq, burada $|W_j|$ W_j vektorunun uzunluğu; $|X|$ giriş vektorunun uzunluğu, $\cos(W_j, X)$ giriş vektoru ilə j -cu çəki vektoru arasındakı bucaqdır. Beləliklə, W_j və X vektorlarının bir-birinə yaxınlaşıması (6.44) ifadəsini maksimallaşdırır, belə ki, kiçik bucaqların kosinusu 1-ə yaxın olan qiymət alır.

Rəqabətli öyrətmə alqoritminin sadə və aydın həndəsi şərhi var. Üç giriş neyronu olan şəbəkəyə baxaq. Əgər hər bir giriş vektoru və çəki vektorunun uzunluğu vahidə bərabərdirsə, onda onları vahid kürənin səthi üzərində nöqtələrlə təsvir etmək olar (şəkil 6.22).

Şəkildə giriş vektorları "+" işarəsi ilə, çəki vektorları isə kiçik çevrələrlə qeyd olunmuşlar. Əgər öyrətmə sürətini kifayət qədər kiçik seçsək, onda çəki əmsalları vektoru uyğun klasterlərin ağırlıq mərkəzinə yaxınlaşan zaman proses stabilləşəcək. Qeyd etdiyimiz kimi, qeyri-supervizor alqoritmlərində neyron şəbəkənin arzuolunan davranışı verilmədiyindən onun həqiqi davranışının meyletməsini tapmaq mümkün deyil. Bu halda öyrətmə xətasını təyin etmək üçün

$$E = \frac{1}{2}\sum_k (w_k^w - x_k)^2$$

münasibətindən istifadə etmək olar, burada w_k^w - qalib neyronun çəki vektorunun k -cı komponentidir.



Şəkil 6.22 Rəqabətli öyrətmənin həndəsi şərhi

Təsdiqetmə ilə öyrətmə alqoritmləri.

Supervizor tərəfindən daxil edilən və şəbəkənin davranışının təqdir edilməsini və ya pislənməsini göstərən bir bit informasiyadan istifadə etməklə şəbəkənin davranışını təşkil edən öyrətmə üsulları təsdiqetmə ilə öyrətmə alqoritmləri (reinforcement learning) adlanır. Özünün inkişafı gedişində, uşaq ona tanış olmayan xarici aləmdə öz davranışını təşkil edə bilər. Əgər uşağın etdiyi hərəkətlər onun üçün arzu olunmaz aqibətə səbəb olursa (ağrı, böyüklər tərəfindən pislənmə və s.), onda aydındır ki, uşaq gələcəkdə belə hərəkətlərdən çəkinəcəkdir. Məsələn, uşaq qaynar əşyalara əl vurduqda əli yanır, ağrı ilə cəzalanır. Yaxud da, o ayaqqabılarını tez-tez eyvandan tullayırsa, buna görə də o valideynləri tərəfindən cəzalana bilər. Yəqin ki, bunları bilən uşaq çalışacaq ki, bütün qaynar şeylərə toxunmasın, şeyləri atmasın. Uşağın hansı hərəkətləri mühit tərəfindən təqdir edilsə (məsələn, böyüklərin onu tərifləməsi) onda o bu hərəkətləri möhkəmləndirməyə çalışacaq.

Neyron şəbəkələrini təsdiqetmə ilə öyrətmə alqoritmləri uşağın göstərilən inkişaf məqamlarını xatırladır. Onlar bioloji həqiqətə bənzər sayılır. Bu alqoritmlərdə, qeyd edildiyi kimi, bir yeganə siqnal, təsdiq (təkzib) və ya təqdir (cəza) siqnalı istifadə olunur (*reward/penalty*). Şəbəkə təsdiqetmə siqnallarını aldıqda cari davranışını "möhkəmləndirməyə" çalışır ki, oxşar hallarda onu müvəffəqiyyətlə təkrar edə bilsin. Təkzib siqnalı daxil olan zaman çəki əmsalları elə dəyişdirilir ki, şəbəkənin gələcək davranışı cari davranışdan uzaq olsun.

L_{r-p} ***və*** A_{r-p} ***alqoritmləri.*** Öyrətmədə təsdiq etmə/təkzib etmə siqnalından istifadə edilməsi avtomatlar nəzəriyyəsindən məlumdur [17,12]. Uyğun ehtimallar əsasında həyata keçirilən hərəkətlərdən birini seçən avtomata baxaq. Hərəkəti seçəndə mühit ("müəllim") seçilmiş hərəkətə öz münasibətini (cavabını) bildirir. Bu münasibət özünü "uğur" (təsdiqetmə, mükafatlandırma) və ya "uğursuzluq" (təkzib etmə, cəzalandırma) siqnalı kimi göstərir. Bu əks əlaqəyə əsaslanaraq, avtomat uyğun ehtimalları elə dəyişir ki, təsdiq etmə siqnalının gözlənilməsi artsın. Təsdiq etmə siqnallarının verilmə tezliyi lazımi yüksək səviyyəyə çatana qədər bu proses davam edir.

Bu sinifdən olan alqoritmlərdən birinə, L_{r-p} mükafatlandırma/cəzalandırma xətti alqoritminə baxaq (*Linear reward/penalty algorithm*) [17]. Avtomatın seçdiyi hərəkətlər çoxluğunu ($a^1, a^2, \ldots, a^r$), uyğun ehtimallarını isə ($p^1, p^2, \ldots, p^r$) ilə işarə edək. Öyrətmənin gedişində hərəkətlərin seçilməsi ehtimalları dəyişdiyi üçün əlavə k indeksi daxil edək, (p_k^1, p_k^2, ...,p_k^r) k-cı öyrətmə addımında cari ehtimallar çoxluğu olsun. Bu addımda a^i hərəkətini seçdikdə mühit b_k cavabını göndərir. b_k yalnız iki qiymət: -1 (0) və ya +1 ala bilər. Bu siqnaldan başqa heç bir əlavə informasiya verilmir. Onda avtomat p_k^i ehtimallarını aşağıdakı qaydalara uyğun olaraq dəyişir.

Əgər seçilmiş hərəkətlər $a_k = a^i$ və $b_k = 1$ isə, onda

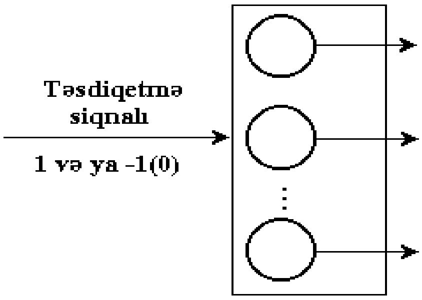
$$\begin{cases} p_{k+1}^i = p_k^i + \gamma_1(1 - p_k^i) \\ p_{k+1}^j = (1-\gamma_1)p_k^j, \quad j \neq i \end{cases} \tag{6.45}$$

əgər $b_k = -1$ olarsa, onda

$$\begin{cases} p_{k+1}^i = (1-\gamma_2)p_k^i \\ p_{k+1}^j = \dfrac{\gamma_2}{r-1} + (1-\gamma_2)p_k^j, \ j \neq i. \end{cases} \tag{6.46}$$

Burada γ_1 və γ_2 öyrətmə sürətidir. Beləliklə, uğurlu hərəkətin ehtimalı vahidlə k-cı addıma qədər olan ehtimalın fərqinə mütənasib kəmiyyət qədər artır. Bu zaman o biri hərəkətlərin ehtimalları azalır. Aydındır ki, γ_1 və γ_2 0 və 1 arasındadır.

Neyron şəbəkələrə keçərək, şəkil 6.23-də təsvir olunmuş neyronlar çoxluğuna baxaq. a^i hərəkəti əvəzinə i-ci neyronun fəallaşmasına baxacağıq. Avtomatda olduğu kimi, burada da öyrətmə elementlərin işədüşmə ehtimallarının dəyişməsi hesabına aparılır.

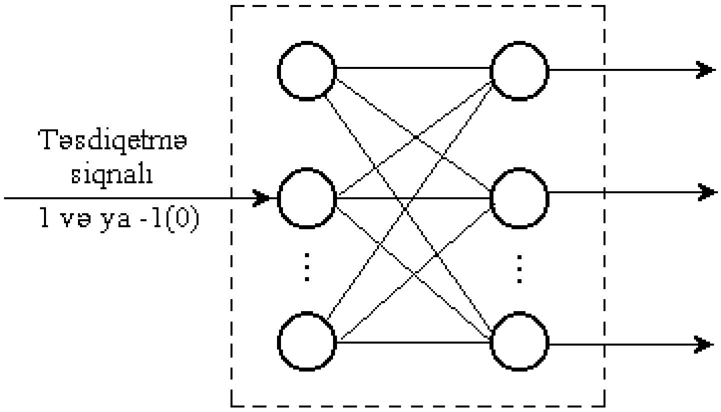


Şəkil 6.23 L_{r-p} ilə öyrədilən neyronlar

Lakin bu halda, ayrılıqda fəaliyyət göstərən neyronlar şəbəkə təşkil etmirlər və əlaqələrdən istifadə etmək hesabına alınmış gücdən məhrumdurlar, assosiasiya və təsnifat apara bilmirlər.

Yuxarıda göstərilən səbəbdən Barto və Anandan [20] genişlənmiş L_{r-p} alqoritmi işləmişlər. Onlar bu alqoritmi "təsdiqetmə ilə assosiativ öyrətmə" (*associative reinforcement learning*) adlandırmışlar.

Onlar həm çıxış, həm də giriş elementləri olan şəbəkəyə baxmışlar (şəkil 6.24).



Şəkil 6.24 A_{r-p} alqoritmi üçün şəbəkə

Təsnifat aparmaq məqsədi ilə şəbəkənin girişinə vektorlar verilir. Korrekt təsnifat aparmaq üçün şəbəkə müəllimin verdiyi qlobal təsdiqetmə/təkzibetmə siqnallarından istifadə edilməklə öyrədilir. Şəbəkəyə onun davranışının səhihliyini bildirmək üçün müəllimin X_k giriş vektorları ilə Y_k çıxış vektorları arasındakı uyğunluq haqqında informasiyası olmalıdır. Bu informasiya $d(X_k,\ Y_k)$ massivi şəklində saxlanıla bilər.

Şəbəkənin çıxışında iki neyron varsa, yəni giriş vektorları iki sinifə ayrılacaqsa, onda X giriş vektoru verildikdə $p(y_1 \mid X) > p(y_2 \mid X)$ şərti ödəndikdə y_1 çıxışı fəallaşırsa və $p(y_2 \mid X) > p(y_1 \mid X)$ şərti ödəndikdə y_2 çıxışı fəallaşırsa, təsnifat xətalarının sayı azalır. Lakin buradakı şərti ehtimalların təyin edilməsi problemi meydana çıxır. Barto və Anandan bu ehtimalları aproksimasiya edən $\boldsymbol{\theta}$ vektorundan istifadə etməyi təklif etmişlər.

$$\boldsymbol{\theta} X = p(y_1 / X) = p(y_2 / X)$$

Beləliklə, $\boldsymbol{\theta} X > 0$ şərti ödənildikdə y_1 vektoru aktivləşir, $\boldsymbol{\theta} X < 0$ olduqda isə ikinci neyron işə düşür. $\boldsymbol{\theta}$ vektoru öyrətmə gedişində sazlanır.

Bundan əlavə, giriş vektorunun sinifini göstərmək üçün z nişanı daxil edilir. Əgər X vektoru y_1-ə uyğun sinifdə yerləşərsə, onda $z = 1$, əks halda $z = -1$. İsbat etmək olar ki,

$(\boldsymbol{\theta} \cdot X - z)^2$ kəmiyyətinin riyazi gözləməsinin minimallaşdırılması təsnifat xətalarının minimallaşdırılmasına gətirir.

Barto və Anandan Robinson-Monro alqoritmindən istifadə etmişlər [19]. E xətasının $\boldsymbol{\theta}$ -ya görə xüsusi törəməsi

$$\frac{\partial E}{\partial \boldsymbol{\theta}} = -(\boldsymbol{\theta} X - z)X \tag{6.47}$$

kimi təyin edilir.

(6.47) münasibəti $\boldsymbol{\theta}$ vektorunu öyrətmə gedişində sazlamaq üçün istifadə olunur.

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \rho_k(\boldsymbol{\theta}_k X_k - Z_k)X_k \tag{6.48}$$

burada $\boldsymbol{\theta}_k$ müxtəlif öyrətmə addımlarında qiymətləri müxtəlif olan sabitlərdir. Öyrətmə prosesində onlar kiçilirlər və prosesin yığılmasına təsir edirlər.

$\boldsymbol{\theta}$ vektorunun komponentlərinə giriş elementlərini çıxış elementlərindən biri ilə birləşdirən əlaqələrin çəki əmsalı kimi baxmaq olar. Bu çıxış neyronları ümumi əmsallaşdırılmış müsbət giriş daxil olanda fəallaşır və əks halda isə fəallaşmır.

A_{r-p} mükafatlandırma/cəzalandırma assosiativ alqoritmi işlənilib hazırlanan zaman təsadüflik elementindən istifadə edilmişdir. Barto və Anandan hər bir çıxış neyronunun iki vəziyyətdə: 1 və –1 vəziyyətlərində ola bilməsini qəbul etmişlər. Onların işində fəallaşma qaydası

$$\begin{cases} y_k = 1 \ , \ \text{if } \ \boldsymbol{\theta}_k X_k + \xi_k > 0 \ , \\ y_k \quad = \quad - \quad 2 \ , \text{else} \end{cases}$$

şəklindədir, burada ξ_k – paylanması məlum olan halda təsadüfi kəmiyyətdir. X_k və $\boldsymbol{\theta}_k$ -nın verilmiş qiymətlərində $E(y_k / \boldsymbol{\theta}_k, X_k)$ riyazi gözləməsi məlum olur. $\boldsymbol{\theta}_k$ -nın yeniləşməsi Robinson-Monro alqoritmində olduğu kimi, yəni (6.48)-də olduğu kimi qalır.

Təsdiqetmə ($b = 1$) halını təkzibetmə ($b = -1$) halından fərqləndirmək üçün λ əmsalı daxil edirlər. Təsdiqetmə halında

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \rho_k(E(y_k / \boldsymbol{\theta}_k, X_k) - b_k y_k)X_k$$

alırıq.

Təkzibetmə halında

$$\boldsymbol{\theta}_{k+1} = \boldsymbol{\theta}_k - \lambda\rho_k(E(y_k / \boldsymbol{\theta}_k, X_k) - b_k y_k)X_k$$

alırıq.

$\lambda = 0$ olduqda, Barto və Anandan bu alqoritmi assosiativ mükafatlandırma-fəaliyyətsizlik (*Associative reward-inaction*) A_{R-I} alqoritmi adlandırırlar.

Onlar göstərmişlər ki, aşağıdakı şərtlər ödəndikdə çəki vektoru yığılacaqdır:

1. Giriş vektorları xətti asılı deyil;

2. Bu vektorların hər birinin meydana çıxma ehtimalı sonludur;

3. Təsadüfi kəmiyyətin paylanması kəsilməz və monotondur;

4. ρ_k ardıcıllığı k artdıqda onun hədlərinin 0-a qədər kiçilməsini təmin edən müəyyən şərtləri ödəyir.

Təsdiqetmə ilə öyrətmə alqoritmlərinin əsas çatışmamazlığı onların böyük məsələlərin həllində olduqca səmərəsiz olmasıdır. Böyük şəbəkələrdə yalnız bir qlobal siqnalı əsas tutaraq davranışı sazlamaq çətindir. [6]-da doğru qeyd edilmişdir ki, bu alqoritmlərlə böyük neyron şəbəkəni öyrətmək, hər bir adamın öz iş gününün faydalılığını milli gəlirin artımı üzrə verilənlər əsasında qiymətləndirməsinə oxşayardı.

Digər çatışmamazlıq odur ki, təsdiqetmə siqnalının gözlənilməsinin artırılması lokal optimuma gətirib çıxara bilər. Şəbəkənin bu optimuma yaxınlaşdıran addımların sayını artıranda, o digər mümkün hallər haqqında getdikcə daha az informasiya alır.

Genetik alqoritmlər. Təbiətdə gedən təkamül prosesləri sistemlərin özünütəşkil-etməsinə canlı misaldır. Darvinin inkişafın təkamülü haqqında təlimi canlı orqanizmlərin xarici mühitin dəyişmələrinə uyğunlaşması prosesini aydınlaşdırdı. Canlıların hər hansı bir növünün hər bir inkişaf mərhələsində mühitin cari şəraitinə daha çox uyğunlaşan fərdləri sağ qalır. Demək olar ki, təbiət bu və ya digər fərdin cari şəraitə uyğunlaşması barədə "qərar qəbul edir". Daha çox uyğunlaşmış fərdlər sağ qalır və yeni nəsil törədir. Bununla bərabər öz xüsusiyyətlərini, əlamətlərini genlər vasitəsilə yeni nəslin fərdlərinə ötürür. Beləliklə, yeni nəslin dəyişən mühitdə yaşamalarını (sağ qalmalarını) təmin edən keyfiyyətlər onlara öz valideynlərindən irsən keçir. Hər bir yeni nəsil ümumən əcdadlarına nisbətən daha yüksək dərəcədə uyğunlaşma qabiliyyətinə malik olur. Bu nəsil öz növbəsində yeni nəsil törətdir və s. Bir neçə nəsildən sonra populyasiya fərdləri artıq uzaq əcdadlarına nisbətən daha yüksək uyğunlaşma xüsusiyyətlərinə malik olur. Bu təkamül prosesi müntəzəm davam edir və milyon illər ərzində əcdadlarından güclü fərqlənən növlər yaranır.

Təbiətdə özünütəşkilin gözəlliyi və ahəngi bir çox tədqiqatçıların diqqətini cəlb etmişdir. Bioloji təkamülün analogiyası əsasında "genetik alqoritmlər" adlanan üsullar sinfi işlənmişdir [42,43]. Bu üsullarda fərdlərin bəzi populyasiyalarının yaşamalarına baxılır (bunlar neyron şəbəkədə çəkilər çoxluğu ola bilər). Genetik alqoritmlər populyasiyaları daha çox "uyğunlaşan" nəsil almağa məcbur edir. Ən sadə halda fərdlər ikilik kodlaşdırılmış vektorla təsvir edilir. Bu vektorların komponentlərinin mümkün qiymətləri genin iki müxtəlif variantının analoqudur. Populyasiyada fərdlər xarici mühitə müxtəlif cür uyğunlaşır. Fərdin mühitə nə qədər yaxşı uyğunlaşmasını və ya bu fərdin populyasiyanın inkişafında nə qədər xeyirli olmasını qiymətləndirməkdən ötrü yararlılıq funksiyası adlanan funksiya istifadə edilir. Yeni fərd doğulduqda, onlar özlərində valideynlərinə xas olan xüsusiyyətləri birləşdirirlər, yəni müxtəlif valideynlərdən genetik kodun müxtəlif sahələri köçürülür.

İnkişafın t addımında populyasiyaya baxaq. t nəsli öz yeni nəslini, yəni $t+1$ nəslini törədir. Yeni fərdin yaradılması üsullarından biri aşağıdakından ibarətdir. Tutaq ki, t nəslinin hər bir fərd n-ölçülü ikilik vektordur. Onda, yeni fərdin yaradılmasında komponentlərinin qiymətləri əvvəlcədən məlum olmayan n-ölçülü vektor qurulur. Törənən fərdin i-ci komponentinin qiymətini təyin etmək üçün təsadüfi olaraq t nəslindən bir vektor götürülür və onun i-ci komponenti i-ci mövqeyə köçürülür. Törənən fərdlərin hamısı $t+1$ nəsilində qalmır. Seleksiya prosesi yeni nəslin yaradılması üçün yararlılıq funksiyası əsasında yararlı fərdləri seçir. Yararsız fərdlər atılır. Bu təbii seçmə prosesinə oxşayır: zəif fərdlər məhv olur, daha güclü fərdlər isə sağ qalır.

Əgər valideynlər təsadüfi seçilirsə, onda müxtəlif komponentlər arasında əlaqə haqqında informasiya itir. Bu informasiyanın saxlanılması üçün fərdin iki valideyndən törəməsi üsulundan istifadə etmək məqsədəuyğundur. Burada yenə təbiətlə paralel aparmaq olar, belə ki, heyvanların çoxunda yeni fərdin doğulması prosesi iki valideynin iştirakını tələb edir. Aralarındakı əlaqələr əhəmiyyətli olan komponentləri bir-birinə yaxın yerləşdirməklə komponentlər arasında əlaqələrin saxlanılması tendensiyası gücləndirilir, belə ki, adətən bir valideyndən yaxın komponentlərin qiymətləri köçürülür. Beləliklə, populyasiya komponentlərin böyük olmayan kombinasiyalarının baş verməsi ehtimalını da ayrı-ayrı komponentlərin qiymətlərinin ehtimalı kimi təsvir edə bilər. Genetik alqoritmlərdən fərdlərin uyğunlaşma xüsusiyyətlərinin daha böyük olmayan kombinasiyalarla təyin edildiyi hallarda müvəffəqiyyətlə istifadə etmək olar. Xolland bu cür böyük olmayan kombinasiyaları "böhran qrupları" adlandırmışdır.

Genetik alqoritmləri şərti olaraq o səbəbə görə təsdiqetmə ilə öyrətmə üsullarına aid etmək olar ki, seleksiya prosesində fərdin yararlığının təyin edilməsi təsdiqetmə/təkzibetmə siqnalının verilməsinin analoqu olur, öyrətmənin məqsədi isə populyasiyanın uyğunlaşma qabiliyyətinin artırılmasıdır, bu da təsdiq-etmə siqnalının riyazi gözləməsinin artırılması ilə analojidir.

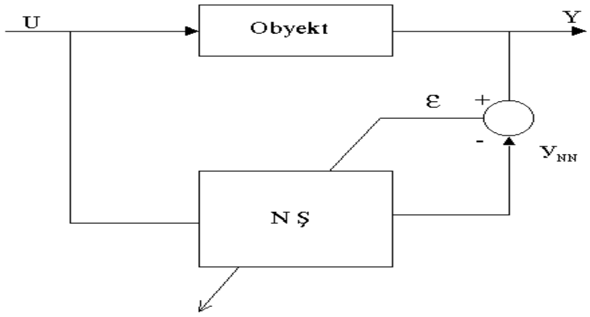
Neyron şəbəkələrin öyrədilməsininin genetik alqoritmlərinə 7-ci fəsildə ətraflı baxılacaq.

6.3 Neyrokompütinqə Əsaslananan İdentifikasiya və İdarəetmə Sistemləri

Neyron identifikasiya. Neyron şəbəkələr funksiyaların universal aproksimatoru kimi mürəkkəb qeyri-xətti naməlum obyektlərin və sistemlərin identifikasiya edilməsində geniş istifadə olunur. Şəkil 6.25-də obyektin düz modelinin neyron identifikasiyasının sxemi təsvir edilmişdir.

Bu halda özünü axtarılan model kimi göstərən neyron şəbəkənin çəkilərinin elə qiymətləri təyin edilir ki, onlar modelin (neyron identifikatorun) Y_{NN} çıxışı və naməlum obyektin cari çıxışı arasında olan xətanın orta kvadratik qiymətinin minimal olması təmin edilir. Bu tip identifikatorda (yəni obyektin birbaşa modelinin identifikatorunda) obyektin çıxışı neyron identifikatorun çıxışı ilə müqayisə edilir.

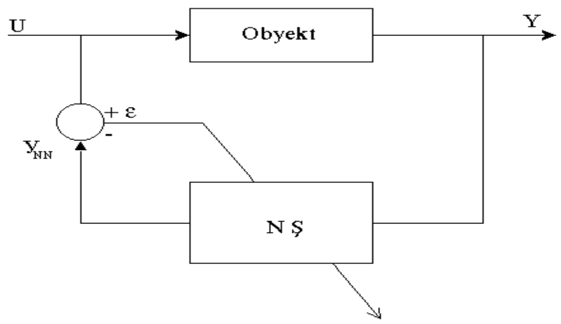
Obyektin tərs modelini təyin edən zaman onun girişi (U) neyron identifikatorun çıxışı (Y_{NN}) ilə müqayisə edilir. İdentifikasiya məsələsi burada belə qoyulur. Obyektin axtarılan modelinin realizə olunduğu neyron şəbəkənin çəki əmsallarının elə qiymətlərini tapmaq lazımdır ki, bu qiymətlərdə ortakvadratik xətanın ε_{ck}^2 qiymətinin minimal olması təmin edilsin (şəkil 6.26).



Şəkil 6.25 Obyektin düz modelinin identifikasiya sxemi

[25-26]-da daha iki növ neyron identifikatoru: adaptiv və öyrətmə identifikatorunu fərqləndirirlər. Birincisində neyron şəbəkənin çıxışı bir sınaq ərzində öyrətmə siqnalına yığılır (şəkil 6.27).

İkinci növ identifikator hər bir sınaq müddətində öyrədilir. Neyron şəbəkənin çıxışı öyrətmə siqnalı ilə bir neçə sınaqdan sonra üst-üstə düşür (şəkil 6.28).



Şəkil 6.26 Obyektin tərs modelinin identifikasiyası

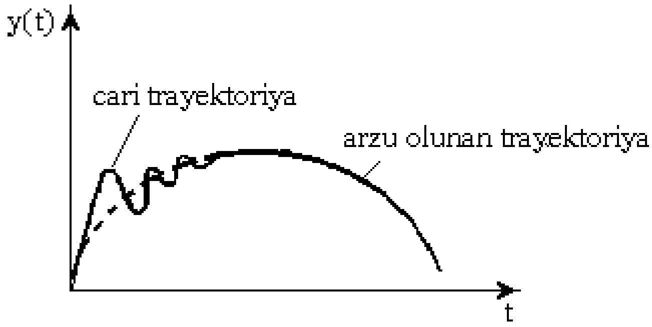
Sadəlik üçün əvvəlcə

$$A(z^{-1})Y(k) = z^{-d}G_0B(z^{-1})U(k)\,,$$

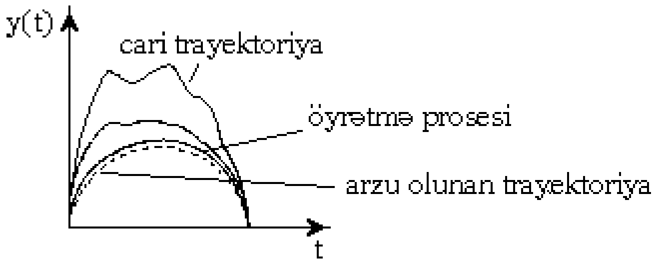
$$A(z^{-1}) = 1 + \sum_{i=1}^{n} a_i z^{-i}\,, \qquad (6.49)$$

$$B(z^{-1}) = 1 + \sum_{i=1}^{m} b_i z^{-i}$$

ötürmə funksiyaları ilə təsvir edilən obyektlər sinfinə baxaq, burada a_i, b_i obyektin naməlum parametrləri, d sırf gecikmə vaxtıdır. Fərz edilir ki, n və m tərtibləri və d sırf gecikmə vaxtı məlumdur.



Şəkil 6.27 Adaptiv növlü identifikasiya



Şəkil 6.28 Öyrətmə növlü identifikasiya

(6.49) obyektinin çıxışı

$$Y(k) = -\sum_{i=1}^{n} a_i Y(k-i) + G_0 \left\{ U(k-d) + \sum_{i=1}^{m} b_i U(k-i-d) \right\} \qquad (6.50)$$

kimi ifadə oluna bilər.

Kitabın 6.2 bölməsində baxılmış öyrətmə üsullarından istifadə etməklə neyron modelin çəki matrislərini elə seçmək olar ki, $\varepsilon^2 = \left[Y(k) - Y_{NN}(k)\right]^2$ minimal olsun. Göründüyü kimi obyektin çıxışı $Y(k)$ düz modelin identifikasiyasında öyrətmə siqnalının yerində durur. (6.50) tənliyini istifadə etməklə neyron identifikatorun $I(k)$ girişi və naməlum parametrlər vektoru α aşağıdakı ifadələrlə təyin oluna bilər [25]

$$I^T(k) = [U(k-d), Y(k-1), \ldots, Y(k-n), U(k-d-1), \ldots U(k-m-d)] \qquad (6.51)$$

$$\alpha^T = [G_0, a_1, \ldots, a_n, G_0 b_1, \ldots, G_0 b_m] \tag{6.52}$$

Öz növbəsində neyron şəbəkənin çıxışı (əgər o xəttidirsə)

$$Y_{NN}(k) = w^T(k)W(k)I(k) \tag{6.53}$$

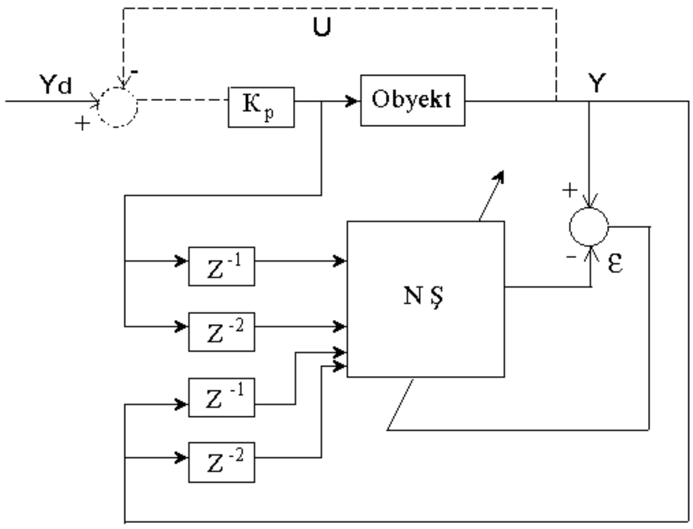
kimi ifadə oluna bilər, burada $w^T(k)$, $W(k)$ çəki vektoru və matrisidir.

Öyrətmə siqnalını $Y(k) = \alpha^T I(k)$ kimi təsvir etsək, onda asanlıqla

$$\varepsilon(k) = [\alpha^T - w^T(k)W(k)]\, I(k) \tag{6.54}$$

alarıq.

(6.54)-ə əsasən müəyyən etmək olar ki, $\varepsilon = 0$ şərti $\alpha^T = w^T(k)W(k)$ deməkdir, bu da $w^T(k)$, $W(k)$ çəki matrisləri və axtarılan α^T parametrləri arasında birbaşa uyğunluq olduğunu göstərir. Şəkil 6.29-da $n = 2$, $m = 1$, $d = 1$ halında obyektin birbaşa ötürmə funksiyasının identifikatorunun quruluşu təsvir edilmişdir. Şəkil 6.29-dan göründüyü kimi, neyron şəbəkənin (identifikatorun) girişləri obyektin giriş və çıxış siqnalları olur. Neyron şəbəkənin girişində öyrətmə vaxtı həmişə məhdud siqnal olması üçün, obyekt mənfi əks əlaqəylə əhatə olunmuşdur. Əgər obyektin çıxışı həmişə məhduddursa, bu əlaqəyə ehtiyac yoxdur.



Şəkil 6.29 Düz modelin identifikatorunun blok-sxemi

Yuxarıda göstərildiyi kimi, naməlum obyektin tərs ötürmə funksiyasını alanda axırıncının giriş siqnalı öyrətmə siqnalı kimi istifadə olunur ki, bu da

$$U(k \ -d) = (1/G_0)Y(k) + \sum_{i=1}^{n} a_i Y(k-i) + G_0 \left\{ U(k-d) - G_0 \sum_{i=1}^{m} b_i U(k-i-d) \right\}$$

kimi təyin edilir.

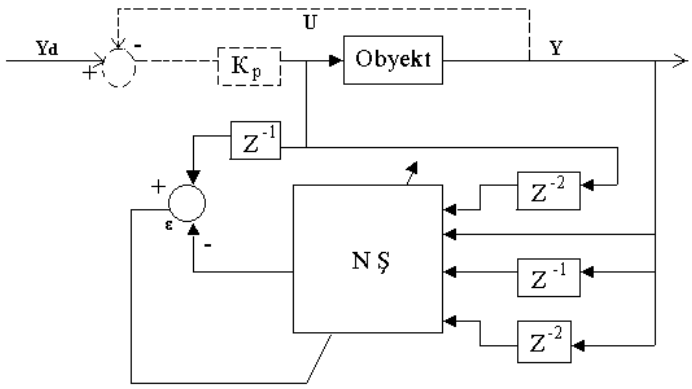
Neyron şəbəkənin giriş vektoru və α naməlum parametrlər vektoru, düz modelin identifikatoruna analoji qaydada təyin edilir

$$I^T(k) = \left[Y(k), Y(k-1), \ldots, Y(k-n), U(k-d-1), \ldots, U(k-m-d) \right] \tag{6.55}$$

$$\alpha^T = \frac{1}{G_0} \left[1, a_1, \ldots, a_n, G_0 b_1, \ldots, G_0 b_m \right] \tag{6.56}$$

$\varepsilon(k) = U(k-d) - U_{NN}(k)$ olduğundan $\alpha^T = w^T(k)W(k)$ olduqda $\varepsilon = 0$ olur. Uyğun öyrətmə alqoritmlərindən istifadə etməklə bu şərt obyektin naməlum parametrlərinin tapılmasına imkan verir.

$n = 2,\ m = 1,\ d = 1$ olan obyekt üçün tərs (invers) neyron identifikatorunun quruluşu şəkil 6.30-da təsvir edilmişdir.



Şəkil 6.30 Düz modelin identifikatorunun blok-sxemi

Biz indiyə kimi xətti neyron şəbəkələri ilə xətti obyektlərin identifikasiyasına baxdıq. Neyron modellərində uyğun qeyri-xətti fəallaşma funksiyası, xüsusi halda, siqmoid funksiyası seçməklə və neyron şəbəkələri vasitəsi ilə istənilən qeyri-xətti obyekti identifikasiya etmək olar.

[25]-də aşağıdakı diferensial tənliklərlə yazılan obyektin düz və tərs modelinin identifikasiyası prosesinin simulyasiyası göstərilmişdir

$$Y(k) = -a_1 Y(k-1) - a_2 Y(k-2) + U(k-1) + bU(k-2) - a_3 Y(k-3) + cY^2(k-1),$$

burada $a_1 = -1.3,\ \ a_2 = 0.3,\ \ b = 0.7$.

Şəkil 6.31-də düz ötürmə funksiyasının identifikatorunun öyrənmə nəticələri təsvir edilmişdir ($a_3 = 0,\ c = 0$). Şəkil 6.31 göstərir ki, düz modelin identifikatoru realizə olunandır.

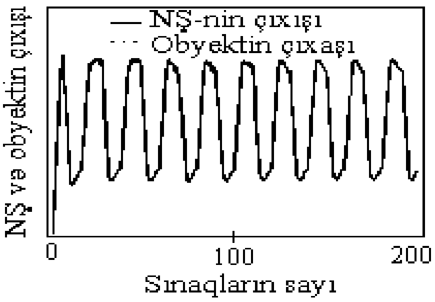
Şəkil 6.32-də qeyri-xətti obyektlər üçün simulyasiyanın nəticələri təsvir edilmişdir ($a_3 = 0.05,\ \ c = 0.1$). İdentifikatoru realizə edən neyron şəbəkəsinin gizli layların neyronlarının fəallaşma funksiyası

$$f(x) = x_g \frac{1 - \exp(-4x / x_g)}{2(1 + \exp(-4x / x_g))}$$

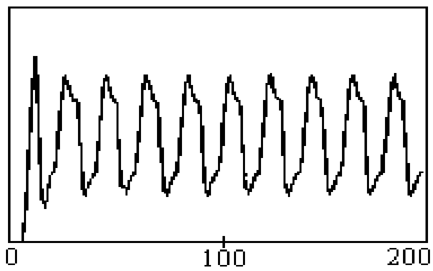
kimi olur.

Şəkil 6.33-də tərs ötürmə funksiyasının identifikatorunun öyrənmə nəticələri təsvir edilmişdir. Şəkil 6.33-dən görünür ki, neyron şəbəkənin çıxışı obyektin girişi ilə, yəni öyrətmə siqnalı ilə üst-üstə düşür. Bu nəticələr tərs ötürmə funksiyasının realizə olunmasınının mümkünlüyünü göstərir.

Neyron idarəetmə. Neyron şəbəkələr müxtəlif idarəetmə sistemlərində tənzimləyici kimi geniş istifadə olunur [27-33]. Bu birinci növbədə onunla əlaqədardır ki, neyron şəbəkələrin quruluşunun çevikliyi qeyri-xətti tənzimləmə qanunlarının geniş spektrini realizə etməyə imkan verir. Bununla da tənzimləyicinin robastlığı artır. İkincisi, öyrənmə qabiliyyəti neyron şəbəkələrə müxtəlif idarəetmə sxemlərini realizə etmək imkanı verir.

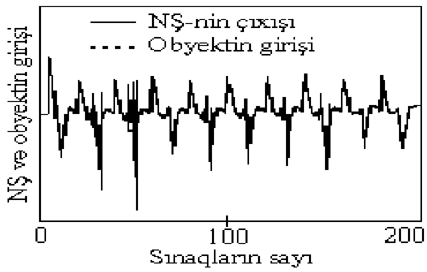


Şəkil 6.31 Düz modelin identifikatorunun öyrənmə prosesi

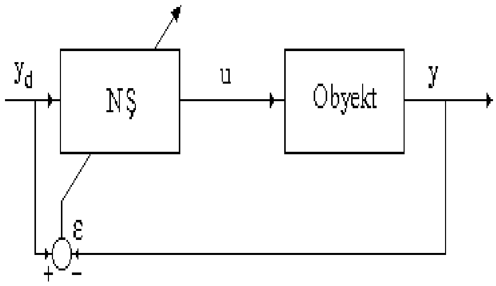


Şəkil 6.32 Qeyri-xətti obyektin düz modelinin identifikatorunun öyrənmə prosesi

Şəkil 6.34-də düz tənzimləyicisi olan neyron idarəetmə sisteminin quruluşu verilmişdir. Bu sistemdə neyron tənzimləyici obyektin tərs modelini realizə edir. Obyektin fiziki realizə oluna bilən tərs modelini həmişə almaq mümkün olmadığına görə neyron idarəetmə sisteminin bu quruluşu həmişə realizə oluna bilmir.



Şəkil 6.33 Tərs modelin identifikatorunun öyrənmə prosesi



Şəkil 6.34 Düz tənzimləyicisi olan neyron idarəetmə sisteminin quruluşu

Şəkil 6.35-də identifikatoru olan neyron idarəetmə sisteminin quruluşu verilmişdir. Bu sinif sistem naməlum xususiyyətləri olan obyektlərin idarə olunması üçün məqsədəuyğundur. Obyektin identifikasiya olunan modeli sistemin neyron tənzimləyiciini sazlamaq üçün istifadə olunur. Neyron identifikatorun belə quruluşuna yuxarıda ətraflı baxılmışdır.

Şəkil 6.36-da üçlaylı neyron şəbəkə əsasında realizə olunmuş xətti neyron tənzimləyicinin quruluşu təsvir edilmişdir.

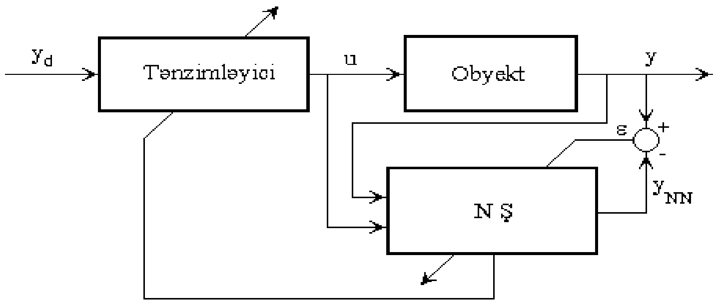
Şəkil 6.37-də qeyri-xətti idarəetmə sistemlərində tənzimləyicini realizə edən üçlaylı neyron şəbəkə göstərilmişdir.

Şəkil 6.38-də texnoloji proses üçün neyron idarəetmə sistemi təsvir edilmişdir.

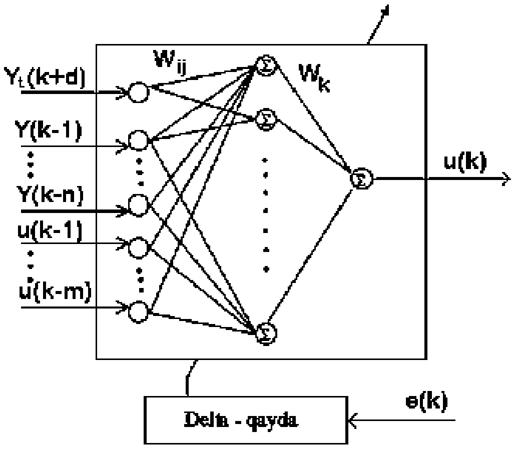
Aşağıda neyron idarəetmə sistemlərinin dayanıqlığı probleminə baxılır [32]. Düz tənzimləyicisi olan neyron sistemin dayanıqlığı araşdırılır. Neyron şəbəkə obyektin dinamikası nəzərə alınmaqla öyrədilir. Orta kvadratik xəta

$$E = e^T e / 2 \tag{6.57}$$

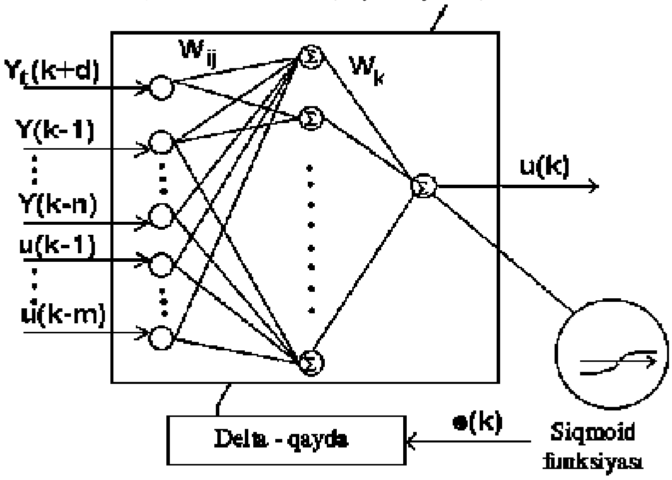
kimi təyin edilir, burada $e = Y - Y_a$.



Şəkil 6.35 İdentifikatorlu sisteminin quruluşu



Şəkil 6.36 Xətti üçlaylı neyron şəbəkə



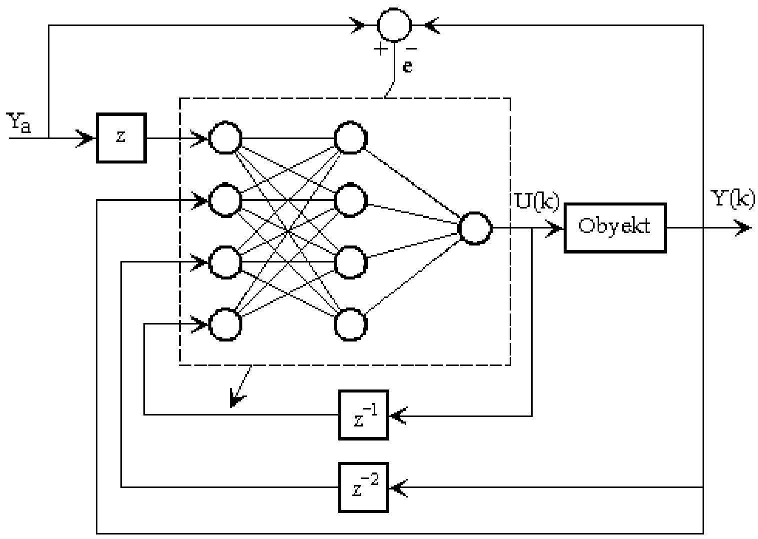
Şəkil 6.37 Qeyri-xətti üçlaylı neyron şəbəkə

Sonra (6.57)-yə əsaslanan xətaların geriyə yayılması (*backpropagation*) üsulu istifadə olunur.

$$\Delta W = \nabla_w E = -e^T (\partial Y / \partial U, \partial U / \partial W) \qquad (6.58)$$

$\partial Y / \partial U$ matrisi obyektin dinamikasını nəzərə alır ki, bu sistemin dayanıqlığının tədqiqində zəruridir.

Sistemin quruluşuna müvafiq olaraq, neyron tənzimləyicinin U çıxışı $U = N(W, I)$ kimi təyin edilir.



Şəkil 6.38 Dinamik neyron idarəetmə sisteminin quruluşu

Əgər bütün neyronlar eyni bir $f(x)=1/(1+e^{-x})$ siqmoid funksiyasına malikdirlərsə, onda

$$N(W,I)=f(W_1 f(W_2)I)$$

ifadəsi alınar, burada W_1 giriş və gizli lay arasında çəki matrisi, W_2 gizli və çıxış layı arasında çəki matrisidir.

W_1 və W_2 çəki matrisləri və W çəki vektoru arasındakı münasibətlər

$$W_1=\begin{bmatrix} W_1^{t_1} \\ \vdots \\ W_1^{t_j} \\ \vdots \\ W_1^{t_n} \end{bmatrix}; \quad W_2=\begin{bmatrix} W_2^{t_1} \\ \vdots \\ W_2^{t_j} \\ \vdots \\ W_2^{t_n} \end{bmatrix}; \quad W^t=\left(\ldots W_1^t j,\ldots,W_2^t i\right)$$

ifadələrinin köməyi ilə verilir.

Lyapunov üsulunu tətbiq edək. Lyapunov funksiyası [34]-də olduğu kimi

$$V(e,w)=e^T Pe/2+W^t\Gamma^{-1}w/2$$

qəbul olunur.

Tarazlıq nöqtəsindən olan sapmalar [32]-də olduğu kimi $e(t)=Y(t)-Y_a(t)$ (çıxışın xətası); $w=W-W_0$ (parametrlərin xətası); $u(t)=U(t)-U_0(t)$ (idarəedici girişin xətası) təyin edilir, burada $Y_a(t)$ çıxışın arzu olunan qiyməti, W_0 çəki vektoru (axtarılan vektor bununla üst-üstə düşür), $U_0(t)$ tarazlıq nöqtəsində neyron şəbəkə tərəfindən yaradılan idarəedici giriş olub $U_0 = N(W_0, I_0)$ kimi verilir, I_0 tarazlıq nöqtəsində neyron şəbəkənin girişidir.

Müsbət həqiqi obyektə vahid giriş siqnalı təsir etdiyi halda aşağıdakı münasibətləri alarıq

$$\dot{e}(t) = Fe(t) + gu(t) \tag{6.59}$$

$$\varepsilon(t) = c^T e(t) \tag{6.60}$$

$$PF + F^T P = -Q \tag{6.61}$$

$$Pg = c \tag{6.62}$$

burada $\dot{e}(t) = d(e(t))/d\,t$, F $n \times m$ ölçülü sabit matris, c və g isə sabit vektorlardır. Əgər obyekt müsbət həqiqi sistemdirsə, onda müsbət müəyyən Q matrisi var.

$(dV/d\,t) < 0$ şərti daxilində həm dinamik sistemin dayanıqlı olmasına, həm də neyron şəbəkənin yığılmasına zəmanət verilir, çünki, Lyapunov funksiyası kimi müsbət müəyyən funksiya iştirak edir. Bu şərt (6.59)-(6.62) ifadələrindən istifadə edilməklə araşdırılır

$$\begin{aligned} dV/dt &= e^T (PF + F^T P)e/2 + (e^T Pg + g^T Pe)u/2 + d(w^T \Gamma^{-1} w/2)/dt = \\ &= -e^T Qe/2 + \varepsilon u + \dot{w}^T \Gamma^{-1} w \ . \end{aligned} \tag{6.63}$$

(6.63)-də birinci həldd mənfi olduğundan $dV/d\,t < 0$ şərti

$$\varepsilon u + \dot{w}^T \Gamma^{-1} w \le 0 \tag{6.64}$$

münasibəti ödəndikdə öyrətmə üçün korrelyasiyanın qiymətini verir, burada

$$\dot{w} = dw/d\,t = \Delta w/\Delta t \ .$$

(6.64) şərtinə bərabərlik kimi baxdıqda, yəni $\varepsilon u + w^T \Gamma^{-1} = 0$ olduqda aşağıdakı öyrətmə qaydası alınır

$$w^T = -\varepsilon u w^{-T} \Gamma (w = \varepsilon u w^{-T}) \tag{6.65}$$

burada w^{-T} vektor kimi təyin edilir və $w^{-T} w = w^T w^{-T} = 1$ şərtini ödəyir.

Xətaların geriyə yayılması üsulu ilə öyrədilən düz tənzimləyicisi olan sistemin dayanıqlığının yoxlanılması aşağıdakına gətirilir. Lyapunov funksiyası kimi həm model üçün, həm də parametrlərin xətası üçün kvadratik forma seçilir. Lakin bu halda, xətaların geriyə yayılması üsulu (6.57) vasitəsilə təyin olunmuş E kvadratik xətadan çıxdığından Lyapunov funksiyası kimi aşağıdakı genişlənmiş v(e) kvadratik xətası seçilir.

$$v(e) = E = e^T Pe/2 \tag{6.66}$$

(6.57) tənliyi (6.66) tənliyinin xüsusi halı kimi verilir. Belə ki, (6.66) tənliyində $P=1$ (vahid matris) götürsək, (6.57) tənliyini alarıq.

(6.59)-(6.62)-dən istifadə edən zaman biz güman edirik ki, obyekt həqiqi müsbət sistemdir və girişi vahid funksiya şəklindədir. Onda dV/dt aşağıdakı kimi alınır.

$$dV/dt = e^T\left(PF + F^T P\right)e/2 + \left(e^T Pg + g^T Pe\right)u/2 =$$
$$= -e^T Qe^T/2 + \varepsilon u = -e^T Qe/2 + \varepsilon N(W,I) \qquad (6.67)$$

Xətaların geriyə yayılması üsulu

$$\Delta W = -\eta e^T P(\partial e/\partial u)(\partial u/\partial W) = -\eta e^T Pg(\partial N/\partial W) = -\eta\varepsilon(\partial N/\partial W) \qquad (6.68)$$

kimi verilir, burada η öyrətmə əmsalının sazlanması üçün müsbət parametrdir. (6.68) ilə təyin edilmiş xətaların geriyə yayılması üsulu (6.67)-də ifadə edilə bilmədiyindən biz tarazlıq nöqtəsindən kiçik meyletmələri nəzərdə tuturuq. Bu lokal dayanıqlığı təhlil etmək zərurəti ilə bağlıdır. İdarəedici u girişi

$$\Delta u = (\partial N/\partial W)^T \Delta W(\partial N/\partial I)\Delta I. \qquad (6.69)$$

kimi verilir.

(6.69) şərtini nəzərə alaraq (6.67)

$$dV/dt = -e^T Qe/2 + \varepsilon\left\{(\partial N/\partial W)^T \Delta W + (\partial N/\partial I)^T \Delta I\right\} =$$
$$= -e^T Qe/2 - \eta\varepsilon^2(\partial N/\partial W)^T(\partial N/\partial W) + \varepsilon(\partial N/\partial I)^T \Delta I \qquad (6.70)$$

kimi ifadə edilə bilər, burada əks əlaqə məlumatı

$$\Delta I^T = (e,u)^T, \quad u = (g/|g|)u$$

kimi təyin edilir.

(6.70) tənliyi təzədən belə yazılır

$$dV/dt = -e^T Qe/2 + \eta\varepsilon^2(\partial N/\partial W)^T(\partial N/\partial W) + \varepsilon(\partial N/\partial e)^T e + \varepsilon(\partial N/\partial u)^T u =$$
$$= -e^T(Q/2 + R - S)e + e^T Tu \qquad (6.71)$$

burada

$$R = \eta Pg(\partial N/\partial W)^T(\partial N/\partial W)(Pg)^T,$$
$$S = Pg(\partial N/\partial e)^T, \qquad (6.72)$$
$$T = Pg(\partial N/\partial u)^T.$$

$dV/dt \le 0$ olduqda dayanıqlığa zəmanət verildiyindən, biz (6.71)-in hər bir həddinin işarəsini araşdırırıq. Lakin ($e^t Tu$) ikinci həddin işarəsi təyin edilməmişdir. $\int(dV/dt)\delta t$ inteqralının işarəsini müəyyən vaxt intervalında qiymətləndirmək olar. Fərz edilir ki, əgər

$$\int_{-\infty}^{m} (e^T u)\, dt \geq 0 \tag{6.73}$$

olarsa, obyekt – müsbət həqiqi sistem passiv sistemə gətirilir [35].

$T < 0$ olduqda ikinci həddin işarəsi mənfi olur. Belə ki, $Q, R > 0$ ümumilikdə ödənilir, $S,T < 0$ olduqda $\int (dV / dt)\delta t$ inteqralının işarəsi mənfi olur. Nəticədə sistemin dayanıqlığı təmin olunur. S və T-nin mənfi müəyyənliyi $(\partial N / \partial I)$-nin qiymətindən asılıdır.

$$\partial N / \partial I = f'(net^o) W_1 f'(net^i) W_2 \tag{6.74}$$

kimi hesablanır, burada

$$f'(net^o) = d(f(W_1 f(W_2 I))) / d(W_1 f(W_2 I)),$$

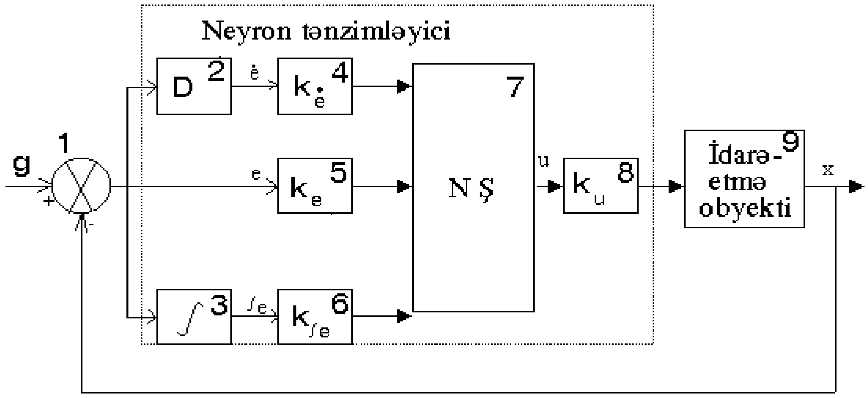
$$f'(net^i) = d(f(W_2 f)) / d(W_2 I).$$

(6.74)-də göstərildiyi kimi, S və T-nin mənfi müəyyənliyi neyron şəbəkənin çəki vektorundan asılıdır. Bu mənfi müəyyənlik çəki vektorunun başlanğıc qiymətindən və yığılma trayektoriyası ilə sıx bağlıdır.

Bu nəticələr göstərir ki, S və T mənfi müəyyənliyi əldə edildikdə sistem dayanıqlı olur. Bununla belə, hətta mənfi müəyyənlik şərtləri ödənilmədikdə belə, müsbət müəyyən Q və R matrislərinin normaları böyük qiymətlər alanda sistem dayanıqlı ola bilər. Q matrisinin müsbət müəyyən olması müsbət həqiqi obyektdən asılıdır. (6.71)-də göstərildiyi kimi, müsbət müəyyən R matrisinin norması həm η sazlama parametrindən, həm də çəki matrisinin normasından çox asılıdır. Buna görə də, η və ($\partial N / \partial W$) norması üçün böyük qiymət seçmək mümkün olduqda sistemin dayanıqlı olması imkanı artır. Bununla belə, diskret xətaların geriyə yayılması üsulundan istifadə ediləndə $\|\Delta W\|$ üçün yuxarı sərhəd mövcud olur, bu da ΔW böyük olduqda dayanıqsızlığa aparır.

Əgər W çəki vektorunun yığılma trayektoriyasını izləmək olsa, onda biz dayanıqlıq şərtini miqdarca qiymətləndirə bilərdik. Aparılan təhlil dayanıqlığı ancaq keyfiyyətcə qiymətləndirir və göstərir ki, xətaların geriyə yayılması üsulundan istifadə olunduqda dayanıqlıq həm çəki vektorunun başlanğıc qiymətindən, həm də η parametrindən asılıdır. Yəni xətaların geriyə yayılması üsulu özü-özlüyündə qlobal asimptotik dayanıqlığa zəmanət verə bilməz. η sazlama parametrini və çəki vektorunun başlanğıc qiymətlərini dəyişməklə sınaq və xəta üsulu ilə dayanıqlıq şərtini tapmaq lazımdır.

Rektifikasiya kalonunun temperaturunun avtomatik tənzimlənməsi üçün neyron tənzimləyicinin sintezi məsələsinə baxaq. Texnoloji prosesin birölçülü avtomatik idarəetmə sisteminin quruluşu şəkil 6.39-da təsvir edilmişdir. Tənzimlənən x koordinatı 1 müqayisə qurğusunda g tapşırığı ilə müqayisə olunur və xəta siqnalı $e = g - x$ paralel olaraq 2 diferensiallayıcısına və 3 inteqrallayıcısına daxil olur. Diferensiallayıcısının çıxış siqnalı $(\dot{e})$, inteqrallayıcının çıxış siqnalı $\left(\int e\right)$ və xəta siqnalı (e) 4-6 elementlərində uyğun olaraq $k_{\dot{e}}, k_{\int e}, k_e$ miqyas əmsallarına vurulur. Sonra onlar 7 neyron şəbəkəsinə daxil olur. Burada u idarəedici siqnalı formalaşır. Bu siqnal 8 elementi ilə miqyaslaşdırılaraq 9 obyektinin icra mexanizminə daxil olur. 4-6 və 8 miqyaslaşdırıcı elementlər neyron tənzimləyicini konkret texnoloji prosesə bağlamağa imkan verir.



Şəkil 6.39 Neyron tənzimləyicinin strukturu

Baxılan neyron AİS-in sintezi məsələsi neyron şəbəkənin quruluşunun, parametrlərinin və $k_{\dot{e}}, k_{\int e}, k_e, k_u$ əmsallarının elə qiymətlərinin təyin edilməsindən ibarətdir ki, sistemin cari və arzu olunan keçid xarakteristikası tələb olunan dəqiqliklə üst-üstə düşsün (şəkil 6.39). Tənzimləyicinin biliklər bazası (BB) neyron şəbəkə əsasında reallaşdırılmışdır. Bununla əlaqədar olaraq arzu olunan BB yaradılması sintez olunan neyron şəbəkənin arxitekturunun və çəki əmsalları matrisinin təyin edilməsinə ekvivalentdir.

İstifadə olunan neyron şəbəkənin arxitekturu üçlaylı "feedforward" quruluşa malikdir. Birinci lay giriş neyronlarından (düyünlərindən), orta lay gizli, axırıncı lay isə çıxış neyronlarından ibarətdir. Neyron şəbəkə əsasında AİS-in sintezi məsələsinin mahiyyəti belədir. Tutaq ki, qurulan sistemin arzu olunan (etalon) davranışını əks etdirən

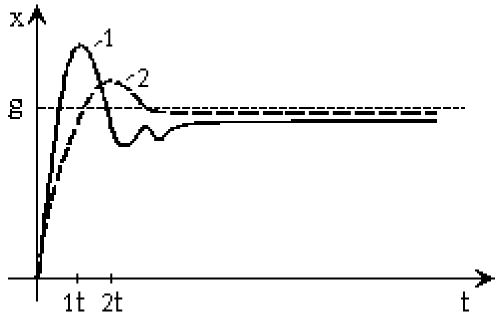
$$\begin{array}{c} \text{IF } g = g_1 \text{ and } x = x_1 \text{ THEN } \dot{x} = \dot{x}_1 \\ \text{IF } g = g_2 \text{ and } x = x_2 \text{ THEN } \dot{x} = \dot{x}_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \text{IF } g = g_n \text{ and } x = x_n \text{ THEN } \dot{x} = \dot{x}_n \end{array} \tag{6.75}$$

produksiya qaydaları toplusu verilib. Layihə olunan AİS-in BB-da bu biliklərin təsvir edilməsi neyron şəbəkənin öyrədilməsindən ibarətdir, yəni neyronların elə çəki əmsalları və hüdudların tapılmalıdır ki, sistemin girişində hər bir cari (g_i, x_i) vəziyyəti əmələ gələndə onun çıxışında (6.75)-ə müvafiq $\dot{x}_i$ cavabları alınsın. Qeyd etmək lazımdır ki, bu məsələnin tanıma məsələsindən mühüm fərqi neyron şəbəkənin öyrədilməsinin qapalı idarəetmə sisteminin tərkibində aparılmasıdır. Yəni neyron şəbəkənin öyrədilməsi üçün tənzimləyicinin çıxışları (yəni neyron şəbəkənin çıxışı) və etalon (obraz) arasında olan xəta deyil, idarəetmə sisteminin arzu olunan xarakteristikası ilə onun cari xarakteristikası arasındakı xəta istifadə olunur.

Neyron şəbəkənin (NŞ) öyrədilməsi xətaların geriyə yayılması alqoritmindən istifadə etməklə aparılır. Şəkil 6.39-a müvafiq olaraq, idarəetmənin xəta siqnalları və onun dəyişmə sürəti (öyrətmə verilənləri) uyğun $k_e, k_{\dot{e}}$ əmsalları ilə miqyaslaşdırıldıqdan sonra neyron şəbəkənin girişinə, yəni sensor neyronlarına verilir.

Yuxarıda şərh olunan yanaşma rektifikasiya kalonunun temperaturunun idarə olunması üçün neyron şəbəkə əsasında qurulmuş AİS-nin sintezində istifadə olunmuşdur. Nəticədə əlaqə çəkilərinin və neyronların hüdudlarının uyğun qiymətləri təyin edilmişdir. Miqyas əmsalları $k_e = 0.5$, $k_{\dot{e}} = 0.05$, $k_u = 0.28$ kimi təyin olunmuşlar. Giriş, gizli və çıxış layında neyronların sayı uyğun olaraq 2, 14, 1-ə bərabərdir. Şəkil 6.40-da ənənəvi PD-tənzimləyicili AİS-nin keçid prosesi (1) və neyron AİS-nin keçid prosesi (2) göstərilmişdir.

Müəyyən edilmişdir ki, ənənəvi PD-tənzimləyicisinin sazlama parametrlərinin optimal qiymətlərində (gücləndirmə əmsalı $K_p = 0.09$ (kqs/sm^2)/dərəcə və diferensiallama sabiti $T_d = 0.15$ dəq.) AİS-də keçid prosesi 18%-li ifrat tənzimləməyə malikdir, tənzimləmə vaxtı $T_p = 1.3 \div 1.5$ dəq., statik xəta $\varepsilon_{st}(\alpha) \approx 0.1x(\alpha)$, inteqral-kvadratik keyfiyyət göstəricisinin qiyməti isə $J = 289.7$. Sistemin statik xətasının bu qiyməti qeyri-qənaətbəxşdir.



Şəkil 6.40 AİS-in keçid proseslərinin əyriləri

Neyron AİS-nin keçid prosesində isə demək olar ki, statik xəta yoxdur ($\varepsilon_{st} \approx 0$), ifrat tənzimləmə 3%-dən çox deyil, tənzimləmə vaxtı $T_p \approx 1.5$ dəq., $J = 160.1$.

Ədəbiyyat

1. Wasserman P.D. Neural Computing: Theory and Practice. Van Nostrand Reinhold, New York, 1989.
2. Zeidenberg M. Neural Networks in Artificial Intelligence. Ellis Horwood Limited, England, 1990.
3. Artificial Neural Networks, Concept Learning, J. Diederich, (Eds.) IEEE Computer Society Press, Los Alamitos GA, 1990.
4. Ackley D.H., Hinton G.E.,Sejnowsky T.J. A Learning Algorithm for Boltzman Machines. Cognitive Science, 9, 1985 pp. 147-169.
5. Kirkpatrick S., Gellat C.D., Vecchi H.P. Optimization by Simulatied Annealing. Science, 220, 1983, pp. 671-690.
6. Hinton G.E. Connectionist Learning Procedures. Artificial Intelligence, vol. 40, N1, 1989, pp. 143-150.
7. Rummelhart D.E., McClelland J.L. (Eds). Parallel Distributed Processing. vol.l: Foundations. The MIT Press, Cambridge, Mass, 1986.
8. Edelman G.E. Neural Darwinism. The Theory of Neuronal Group Selection. Basic Books Inc., Publishers. New York, 1986.
9. Shastri L. Semantic Networks An Evidental Formalization and its Connectionist Realization. Research Notes In Artificial Intelligence. Morgan Kaufman Publ. Inc., San Mateo, California, 1988.
10. Mitchell J.M., Keller R.M., Kedar-Cabelli S.T. Explanation-Based Generalization: A Unifying View. Machine Learning1, 1986, pp. 47-80.
11. Aliev R.A., Vahidov R.M., Aliev R.R.Artificial neural net works theory and practice. Tabriz University, Tabriz-1995.
12. Rummelhart D.E., Hinton G.E., Williams R.J. Learning Internal Representations by Back Propagating errors. Nature 323, 1986, pp. 533-536.
13. Rummelhart D.E., Hinton G.E., Williams R.J. Learning Internal Representations by Errors Propagation, in: D.E. Rummelhart, J.L.McClel-land and the PDR Research Group (Eds.), Parallel Distributed Processing: Explorations in the MicroStructure of Cognition, 1 Foundations.The MIT Press, Cambridge, MA, 1986.
14. Sejnowski T.J., Rosenberg C.R.Parallel Networks that Learn to Pronounce English Text. Complex Systems1, 1987, pp.145-168.
15. Parker O.V. Second Order Back Propogation# emplementing an Optimal 0 (n) approximation to Newton's Method as an Artificial Neural Network. Manuscript Submitted for Publication, 1987.
16. Cooper L.M., Liberman F., Oja E.A. Theory for Application and Loss of Specificity in Visual Contex. Biol. Cybern. 33, 1979, pp. 9-28.
17. Schwartz J.T. The New Connectionism: Developing Relationship Between Neuroscience and Artificial Intelligence, ln:S.R. Craubard (Eds): The Artificial Intelligence Debate: False Starts, Real Foundations, Cambridge. MA, MIT Press.
18. Rummelhart D.E., Zipser D. Competitive Learning. Cognitive Science 9, 1985, pp.75-112.
19. Narenda K.S., Thathachar M.A.L. Learning Automata-a Survey. IEEE Transactions on Systems, Man and Cybernetics,4,1974, pp. 323-324.
20. Barto A.G., Anandan F. Pattern Recognizing Stochastic Learning Automata. IEEE. Transactions on Systems, Man and Cybernetics, 15, 1985, pp. 360-375.

21. Psaltis D., Siderrs A., Yamamura A. Neural Controller in Proc. 1987 IEEE Int. Conf. Neural Networks, vol. IV, San Diego, 1987, pp. 551-558.
22. Guez A., Eibert J., Kam M. Neuromophic architecture adaptive control. A preliminary analysis in Proc. 1987 IEEE Int. Conf. Neural Networks, vol. IV, San Diego, 1987, pp. 567-572.
23. Funahashl K. On the approximatere alization of identity mappings by three-layer neural networks, The Inst. Electron, Inform. Sommun. Eng.,vol..J73-A,no.1, 1990, pp.139-145.
24. Mohamad, Hassoun H. Fundamentals of artificial neural networks. a bradford book. The OMIT press cambridge, Massachusetts, London, England, 1995.
25. Jamada T., Jabuta T. Dynamic system identification using neural networks IEEE Transactions on systems, man, and cybernetics, vol 2, not. January/February 1993.
26. Jabuta T., Jamada T. Possibility of neural network controller for robot manipulators, in proc. Int. Conf. Robotics automat, Cincinnati, OH, 1990, pp. 1686-1691.
27. Narenda K.S., Parthasarathy K., Identification and control of dynamical system using neural networks, IEEE Trans. Neural Networks. vol.11, 1990, pp. 4-27.
28. Miller W.T., Hewes R.P., Glanz F.H., Kraft L.G. Real-time dynamic control of an industrial manipulator using a neural network-based learning controller, IEEE Trans. Robotics Automat, vol 6, 1990, pp. 1-9.
29. Johnson M.A., Leahy M.B., Jr., Adaptive model-based neural network control in Proc. IEEE Int. Conf Robotic Automat., Cincinnati OH, 1990, pp. 1704-1709.
30. Aliev R.R. Neural identification and control of industrial technological processes in Application of Fuzzy Systems edited by R.A.Aliev, Kenarangui R, Tabriz-1994.
31. Aliev R.A., Aliev F.T., Abiev R.H, Aliev R.R. Industrial neural controllers in proc. second European Congress on intelligent Techniques and Soft Computing, Aachen, Germany, 1994, pp. 1321-1326.
32. Yabuta T. and Yamada T. Neural Network controller Characteristics with regard to adaptive control, IEEE Transactions on systems man and cybernetics, vol. 22, no 1, January /February 1992.
33. Jordan M.I. Genetic constraints on underspecified target Otrajectories in Proc. Int Joint Conf. Neural Networks, Washington, DC., vol. 1, 1989, pp. 217-225.
34. Ichikawa K., Adaptive control, Shokoudo, 1984.
35. Craig J.J. Adaptive control of Mechanical Manipulators. Reading, MA: Addison Wesley, 1988.
36. Isabelle Guyon. Neural Networks Systems. Proc. 5-th Int. Symp. Num. Meth. in Eng. vol. 1, Lausanne, 1989.
37. Cohen M.A., Grossberg S.G. Absolute Stability of Global Pattern Formation and Parallel Memory Storage by Sompetitie Neural Networks. IEEE Transactions on Systems, Man and Cybernetics, 1983, pp.815-826.
38. Hopfield J.J. Neural Network and Physical Systems with Emergent Collective Computational Abilities. Proc. of the National Academy of Science 79, 1982, pp. 2554-2559.
39. Hopfield J.J. Neurons with Graded Response Have Collective Computational Properties Like Those of Two-state Neurons. Proc. of the National Academy of Science 81, 1984, pp. 3088-3092.
40. Kosko B. Neural Networks and Fuzzy Systems. A Dynamical System Approach to Machine Intelligence. Prentice-Hall International Inc., 1993, 449 p.
41. Hebb D.O.: The Organization of Behavior. Willey, New York, 1994.
42. Genetic Algorithms and Simulated Annealing, L. Davis (Eds), Pitman, London. 1987.

43. Holland S.H. Adaptation in Natural Artificial Systems, University of Michigan Press, Ann Arbor. MI, 1975.