Дәріс 8

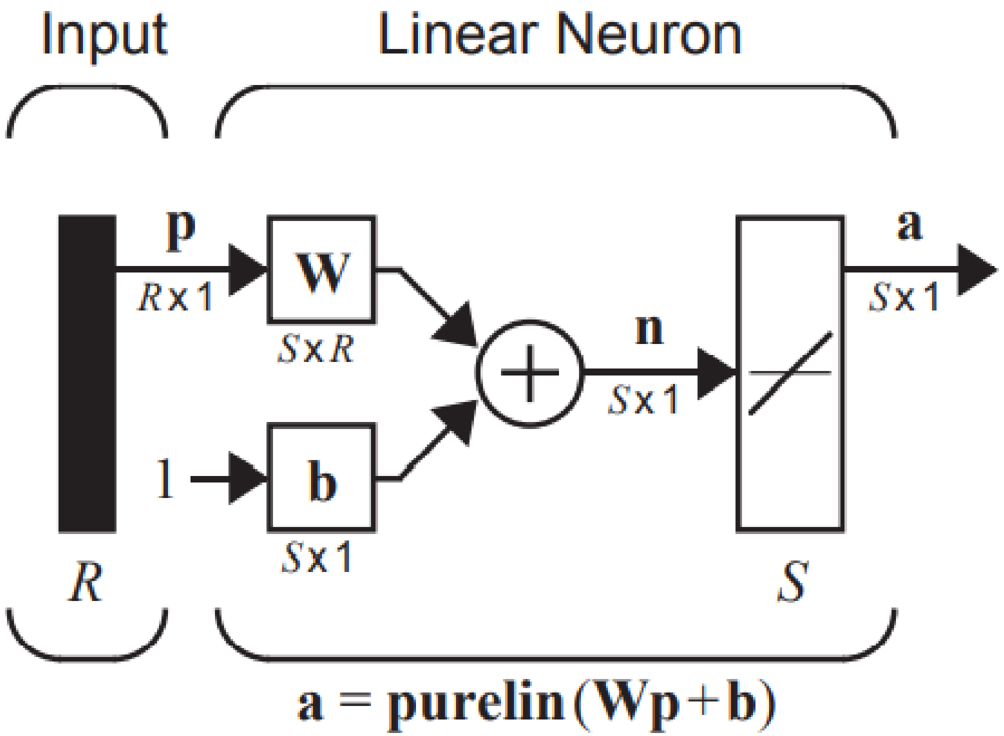
Widrow-Hoff оқытуы

- Алдыңғы дәрістерде біз өнімділікті үйренудің негізін қаладық, онда желі өнімділігін оңтайландыруға үйретіледі. Бұл дәрісте біз бір деңгейлі сызықтық нейрондық желіге өнімділікті үйрену принциптерін қолданамыз.
- Widrow-Hoff оқытуы - өнімділік индексі орташа квадраттық қате болып табылатын шамамен ең тік төмендеу алгоритмі. Бұл алгоритм екі себеп бойынша біздің талқылауымыз үшін маңызды. Біріншіден, ол бүгінде көптеген сигналдарды өңдеу қолданбаларында кеңінен қолданылады, олардың кейбірін осы тарауда қарастырамыз. Сонымен қатар, ол келесі дәрісте берілетін көпқабатты желілер үшін кері таралу алгоритмінің алғышарты болып табылады.

- Бернард Видроу 1950 жылдардың аяғында, Фрэнк Розенблатт перцептронды оқыту ережесін жасаған кезде нейрондық желілерде жұмыс істей бастады. 1960 жылы Уидроу және оның аспиранты Марсиан Хофф ADALINE (ADAptive Linear NEuron) желісін және олар LMS (Ең кіші орташа квадрат) алгоритмі [WiHo60] деп атаған үйрену ережесін ұсынды. Олардың ADALINE желісі перцептронға өте ұқсас, тек оның активациялық функциясы қатаң шектеудің орнына сызықты. ADALINE да, перцептрон да бірдей табиғи шектеуден зардап шегеді: олар тек сызықты түрде бөлінетін есептерді шеше алады. Алайда LMS алгоритмі перцептронды оқыту ережесіне қарағанда күштірек. Перцептрон ережесі жаттығу үлгілерін дұрыс санаттайтын люцияға жақындауға кепілдік берілгенімен, нәтижесінде алынған желілік жұмыс шуға сезімтал болуы мүмкін, өйткені үлгілер жиі шешім шекараларына жақын орналасады. LMS алгоритмі орташа квадраттық қатені азайтады, сондықтан шешім шекараларын оқыту үлгілерінен мүмкіндігінше алыс жылжытуға тырысады.

- LMS алгоритмі перцептронды оқыту ережесіне қарағанда көптеген практикалық қолдануларды тапты. Бұл, әсіресе, цифрлық сигналды өңдеу саласына қатысты. Мысалы, қалааралық телефон желілерінің көпшілігі жаңғырықты жою үшін ADALINE желілерін пайдаланады. Бұл қолданбаларды кейінірек тарауда егжей-тегжейлі қарастырамыз.
- LMS алгоритмінің сигналдарды өңдеу қосымшаларында үлкен табысқа жетуіне байланысты және алгоритмді көпқабатты желілерге бейімдеуде сәттіліктің болмауына байланысты Уидроу 1960 жылдардың басында нейрондық желілердегі жұмысын тоқтатты және адаптивті сигналда толық уақытты жұмыс істей бастады. өңдеу. Ол 1980-жылдары нейрондық желі саласына оралды және өзінің бастапқы LMS алгоритмінің ұрпағы болып табылатын уақытша кері таралуды пайдалана отырып, нейрондық желілерді адаптивті басқаруда пайдалану бойынша зерттеулерді бастады.

ADALINE Network



$$\mathbf{a} = \mathbf{purelin}(\mathbf{Wp} + \mathbf{b}) = \mathbf{Wp} + \mathbf{b} .$$

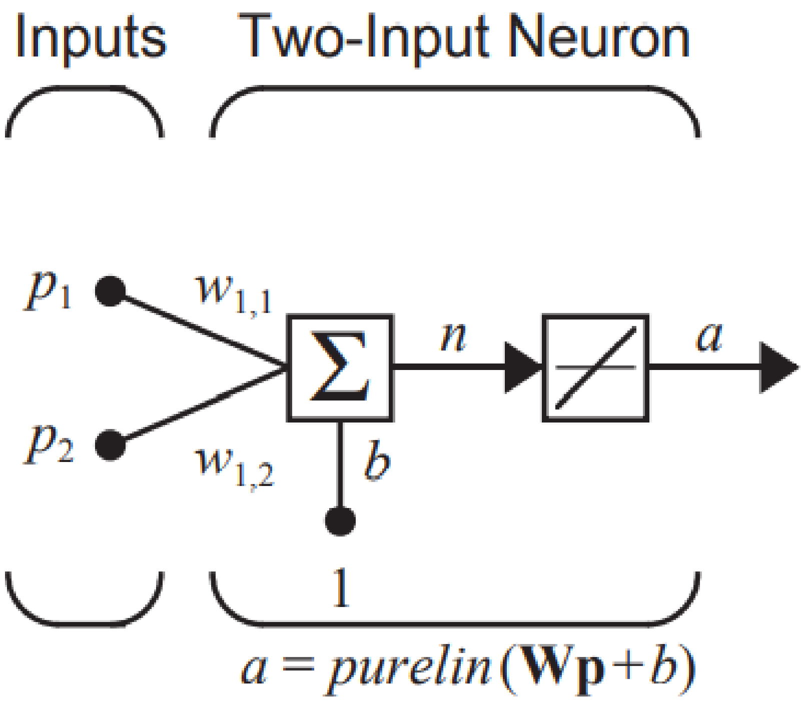
$$a_i = purelin(n_i) = purelin({}_i\mathbf{w}^T\mathbf{p} + b_i) = {}_i\mathbf{w}^T\mathbf{p} + b_i ,$$

$${}_i\mathbf{w} = \begin{bmatrix} w_{i,1} \\ w_{i,2} \\ \vdots \\ w_{i,R} \end{bmatrix} .$$

Бірыңғай ADALINE

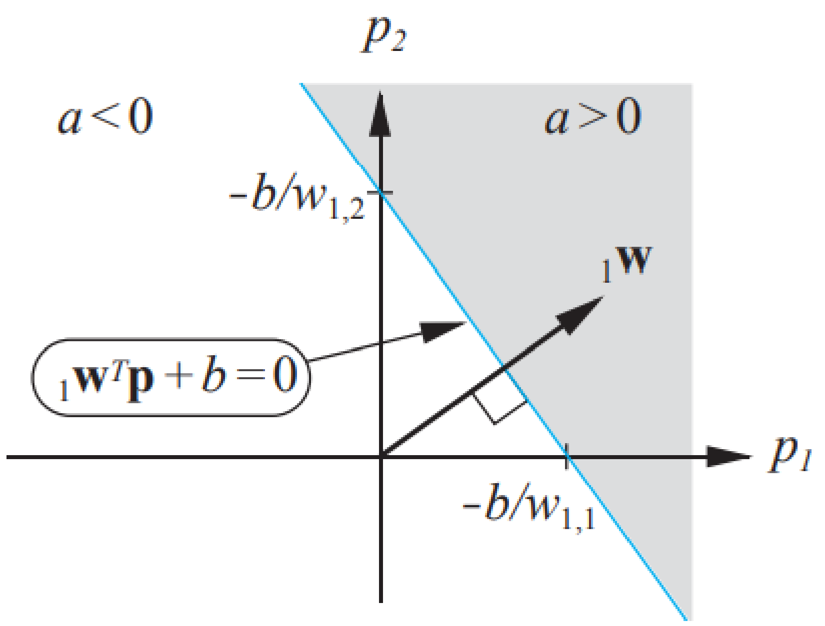
$$a = purelin(n) = purelin({}_1\mathbf{w}^T\mathbf{p} + b) = {}_1\mathbf{w}^T\mathbf{p} + b$$

$$= {}_1\mathbf{w}^T\mathbf{p} + b = w_{1,1}p_1 + w_{1,2}p_2 + b\ .$$



Нейронның шығуы сұр аймақта 0-ден жоғары. Ақ аймақта шығыс нөлден аз. Енді бұл ADALINE туралы нені білдіреді? Онда ADALINE нысандарды екі санатқа жіктеу үшін пайдалануға болады деген сөз бе?

Дегенмен, ол объектілер сызықты түрде бөлінетін болса ғана қолдана аламыз. Осылайша, бұл жағынан ADALINE перцептронмен бірдей шектеуге ие.



Орташа квадрат қатесі

- Енді біз ADALINE желісінің сипаттамаларын зерттегеннен кейін, біз LMS алгоритмін әзірлеуді бастауға дайынбыз. Перцептрон ережесі сияқты, LMS алгоритмі бақыланатын оқытудың мысалы болып табылады, онда оқыту ережесі дұрыс желілік сипаты мысалдарының жиынтығымен қамтамасыз етіледі:

$$\{\mathbf{p}_1, \mathbf{t}_1\}, \{\mathbf{p}_2, \mathbf{t}_2\}, \ldots, \{\mathbf{p}_Q, \mathbf{t}_Q\},$$

- LMS алгоритмі орташа квадраттық қатені азайту үшін ADALINE салмағы мен қиғаштықтарын реттейді, мұнда қате мақсатты шығыс пен желі шығысының арасындағы айырмашылық болып табылады. Бұл бөлімде біз осы өнімділік көрсеткішін талқылағымыз келеді. Алдымен бір нейрондық жағдайды қарастырамыз.

$$\mathbf{x} = \begin{bmatrix} {}_1\mathbf{w} \\ b \end{bmatrix}. \qquad \mathbf{z} = \begin{bmatrix} \mathbf{p} \\ 1 \end{bmatrix}. \qquad a = {}_1\mathbf{w}^T \mathbf{p} + b, \qquad a = \mathbf{x}^T \mathbf{z}.$$

ADALINE желісі орташа квадрат қатесі:

$$F(\mathbf{x}) = E[e^2] = E[(t-a)^2] = E[(t-\mathbf{x}^T\mathbf{z})^2],$$

$$F(\mathbf{x}) = E[t^2 - 2t\mathbf{x}^T\mathbf{z} + \mathbf{x}^T\mathbf{z}\mathbf{z}^T\mathbf{x}]$$

$$= E[t^2] - 2\mathbf{x}^T E[t\mathbf{z}] + \mathbf{x}^T E[\mathbf{z}\mathbf{z}^T]\mathbf{x}.$$

$$c = E[t^2],\ \mathbf{h} = E[t\mathbf{z}] \text{ and } \mathbf{R} = E[\mathbf{z}\mathbf{z}^T].$$

$$F(\mathbf{x}) = c - 2\mathbf{x}^T\mathbf{h} + \mathbf{x}^T\mathbf{R}\mathbf{x},$$

Мұнда **h** вектор кіріс векторы мен онымен байланысты мақсат арасындағы айқаспалы корреляцияны береді, ал **R** кіріс корреляциялық матрицасы болып табылады. Бұл матрицаның диагональ элементтері кіріс векторларының элементтерінің орташа квадрат мәндеріне тең.

$$F(\mathbf{x}) = c + \mathbf{d}^T\mathbf{x} + \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x}.$$

$$\mathbf{d} = -2\mathbf{h} \text{ and } \mathbf{A} = 2\mathbf{R}.$$

Бұл өте маңызды нәтиже, өйткені біз алдыңғы дәрістен квадраттық функцияның сипаттамалары ең алдымен Гессиан матрицасына тәуелді екенін білеміз. Мысалы, егер Гессианның меншікті мәндерінің барлығы оң болса, онда функцияның бір жалпылама минимумы болады. Бұл жағдайда Гессиан матрицасы корреляциялық матрицадан екі есе болады және барлық корреляциялық матрицалар оң анықталған немесе оң жартылай анықталған болып табылатынын көрсетуге болады, бұл олардың ешқашан иелік меншікті мәндерге ие бола алмайтынын білдіреді. Бізде екі мүмкіндік қалды. Корреляциялық матрицада тек оң меншікті мәндер болса, өнімділік индексі бір бірегей жалпылама минимумға ие болады. Егер корреляциялық матрицаның кейбір нөлдік меншікті мәндері болса, өнімділік индексі векторға байланысты әлсіз минимумға ие болады немесе минимум болмайды. $\mathbf{d} = -2\mathbf{h}$

Енді өнімділік индексінің стационарлық нүктесін табайық. Квадраттық функциялар туралы алдыңғы талқылаудан біз градиент екенін білеміз

$$\nabla F(\mathbf{x}) = \nabla\left(c + \mathbf{d}^T\mathbf{x} + \frac{1}{2}\mathbf{x}^T\mathbf{A}\mathbf{x}\right) = \mathbf{d} + \mathbf{A}\mathbf{x} = -2\mathbf{h} + 2\mathbf{R}\mathbf{x}.$$

Стационар нүктесін градиентті нөлге теңестіру арқылы табуға болады

$$-2\mathbf{h} + 2\mathbf{R}\mathbf{x} = 0.$$

Демек, егер корреляция матрицасы оң анықталған болса, күшті минимум болатын бірегей стационарлық нүкте болады:

$$\mathbf{x}^* = \mathbf{R}^{-1}\mathbf{h}.$$

Бұл жерде бірегей шешімнің болуы тек корреляциялық матрицаға байланысты екенін атап өткен жөн. Сондықтан кіріс векторларының сипаттамалары бірегей шешімнің бар-жоғын анықтайды.

LMS алгоритмі

- Видроу мен Хоффтың негізгі түсінігі олар орташа квадраттық қатені бағалай алады

$$\hat{F}(\mathbf{x}) = (t(k) - a(k))^2 = e^2(k),$$

$$\hat{\nabla} F(\mathbf{x}) = \nabla e^2(k).$$

$$[\nabla e^2(k)]_j = \frac{\partial e^2(k)}{\partial w_{1,j}} = 2e(k)\frac{\partial e(k)}{\partial w_{1,j}} \text{ for } j = 1, 2, \ldots, R,$$

$$[\nabla e^2(k)]_{R+1} = \frac{\partial e^2(k)}{\partial b} = 2e(k)\frac{\partial e(k)}{\partial b}.$$

$$\frac{\partial e(k)}{\partial w_{1,j}} = \frac{\partial[t(k) - a(k)]}{\partial w_{1,j}} = \frac{\partial}{\partial w_{1,j}}[t(k) - ({}_1\mathbf{w}^T\mathbf{p}(k) + b)]$$

$$= \frac{\partial}{\partial w_{1,j}}\left[t(k) - \left(\sum_{i=1}^{R} w_{1,i}p_i(k) + b\right)\right]$$

$$\frac{\partial e(k)}{\partial w_{1,j}} = -p_j(k) . \qquad \frac{\partial e(k)}{\partial b} = -1 .$$

$$\hat{\nabla}F(\mathbf{x}) = \nabla e^2(k) = -2e(k)\mathbf{z}(k) . \qquad \mathbf{x}_{k+1} = \mathbf{x}_k - \alpha\nabla F(\mathbf{x})\big|_{\mathbf{x} = \mathbf{x}_k} .$$

$$\mathbf{x}_{k+1} = \mathbf{x}_k + 2\alpha e(k)\mathbf{z}(k) ,$$

$${}_1\mathbf{w}(k+1) = {}_1\mathbf{w}(k) + 2\alpha e(k)\mathbf{p}(k) , \qquad b(k+1) = b(k) + 2\alpha e(k) .$$

LMS алгоритмін матрицалық белгілерде ыңғайлы түрде жазуға болады:

$$\mathbf{W}(k+1) = \mathbf{W}(k) + 2\alpha\mathbf{e}(k)\mathbf{p}^T(k),$$

$$\mathbf{b}(k+1) = \mathbf{b}(k) + 2\alpha\mathbf{e}(k).$$

Мысалы,

- ADALINE желісін және LMS алгоритмін тексеру бастапқыда талқыланған алма/апельсинді тану мәселесін қайта қарастырыңыз.

$$\mathbf{W}(k+1) = \mathbf{W}(k) + 2\alpha e(k)\mathbf{p}^T(k).$$

$$\left\{\mathbf{p}_1 = \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}, t_1 = \begin{bmatrix} -1 \end{bmatrix}\right\} \qquad \left\{\mathbf{p}_2 = \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, t_2 = \begin{bmatrix} 1 \end{bmatrix}\right\}.$$

Егер кіріс векторлары бірдей ықтималдықпен кездейсоқ түрде құрылған деп есептесек, кіріс корреляциялық матрицаны есептей аламыз:

$$\mathbf{R} = E[\mathbf{p}\mathbf{p}^T] = \frac{1}{2}\mathbf{p}_1\mathbf{p}_1^T + \frac{1}{2}\mathbf{p}_2\mathbf{p}_2^T$$

$$= \frac{1}{2}\begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}\begin{bmatrix} 1 & -1 & -1 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}\begin{bmatrix} 1 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix}.$$

$$\lambda_1 = 1.0, \qquad \lambda_2 = 0.0, \qquad \lambda_3 = 2.0 .$$

$$\alpha < \frac{1}{\lambda_{max}} = \frac{1}{2.0} = 0.5 .$$

$$a(0) = \mathbf{W}(0)\mathbf{p}(0) = \mathbf{W}(0)\mathbf{p}_1 = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} = 0 , \qquad e(0) = t(0) - a(0) = t_1 - a(0) = -1 - 0 = -1 .$$

$$\mathbf{W}(1) = \mathbf{W}(0) + 2\alpha e(0)\mathbf{p}^T(0)$$

$$= \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} + 2(0.2)(-1) \begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix}^T = \begin{bmatrix} -0.4 & 0.4 & 0.4 \end{bmatrix} .$$

$$a(1) = \mathbf{W}(1)\mathbf{p}(1) = \mathbf{W}(1)\mathbf{p}_2 = \begin{bmatrix} -0.4 & 0.4 & 0.4 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix} = -0.4 ,$$

$$e(1) = t(1) - a(1) = t_2 - a(1) = 1 - (-0.4) = 1.4 .$$

$$\mathbf{W}(2) = \mathbf{W}(1) + 2\alpha e(1)\mathbf{p}^T(1)$$

$$= \begin{bmatrix} -0.4 & 0.4 & 0.4 \end{bmatrix} + 2(0.2)(1.4)\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}^T = \begin{bmatrix} 0.16 & 0.96 & -0.16 \end{bmatrix} .$$

$$a(2) = \mathbf{W}(2)\mathbf{p}(2) = \mathbf{W}(2)\mathbf{p}_1 = \begin{bmatrix} 0.16 & 0.96 & -0.16 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \\ -1 \end{bmatrix} = -0.64 .$$

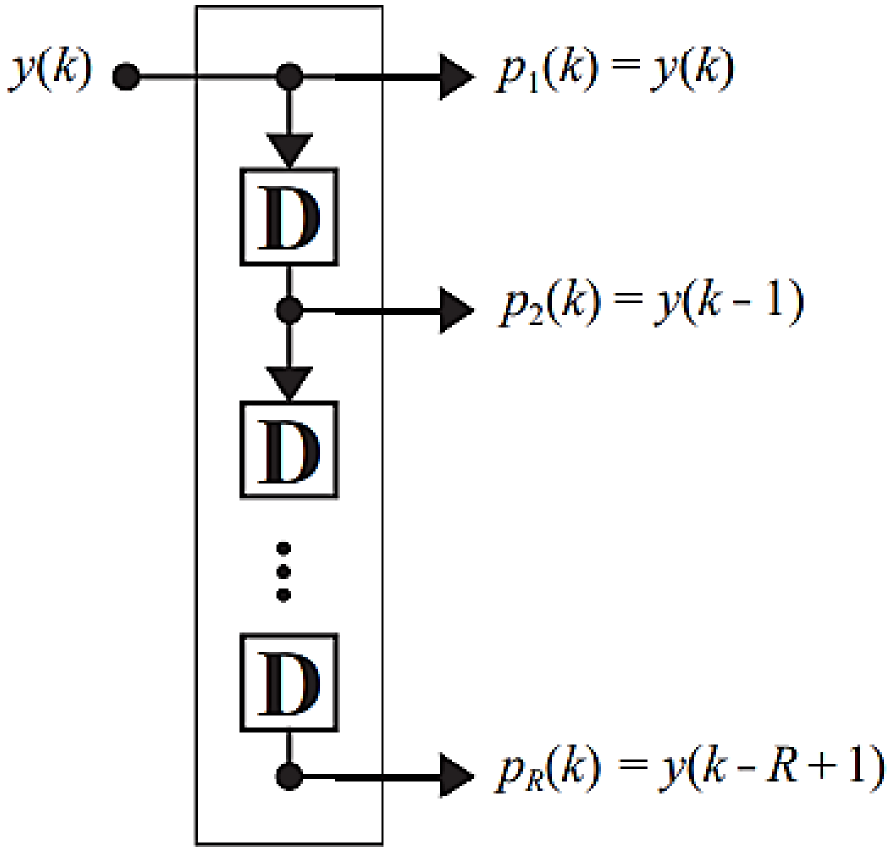
$$e(2) = t(2) - a(2) = t_1 - a(2) = -1 - (-0.64) = -0.36 .$$

$$\mathbf{W}(3) = \mathbf{W}(2) + 2\alpha e(2)\mathbf{p}^T(2) = \begin{bmatrix} 0.016 & 1.1040 & -0.0160 \end{bmatrix} .$$

$$\mathbf{W}(\infty) = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix} .$$

Адаптивті сүзгілеу

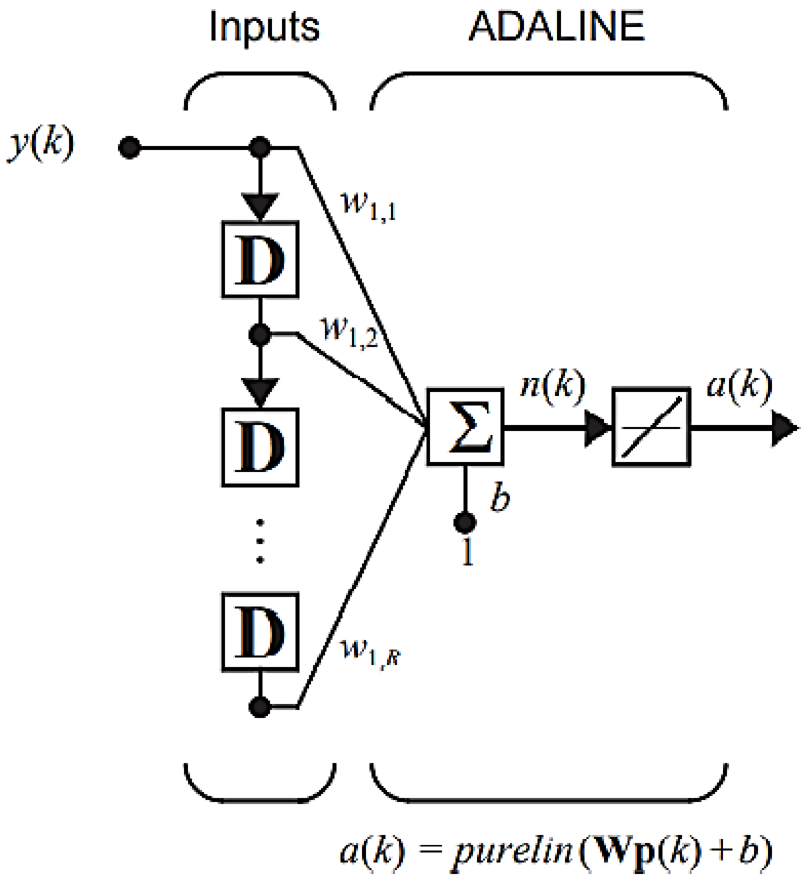
- Осы дәріс басында айтқанымыздай, ADALINE желісі перцептрондық желі сияқты негізгі шектеулерге ие; ол тек сызықты түрде бөлінетін есептерді шеше алады. Осыған қарамастан, ADALINE перцептрондық желіге қарағанда әлдеқайда кеңірек қолданылды. Шындығында, бұл практикалық қолдануда ең көп қолданылатын нейрондық желілердің бірі деп айтуға болады. ADALINE қолданбасының негізгі салаларының бірі адаптивті фильтрлеу болды, онда ол әлі де кеңінен қолданылады.



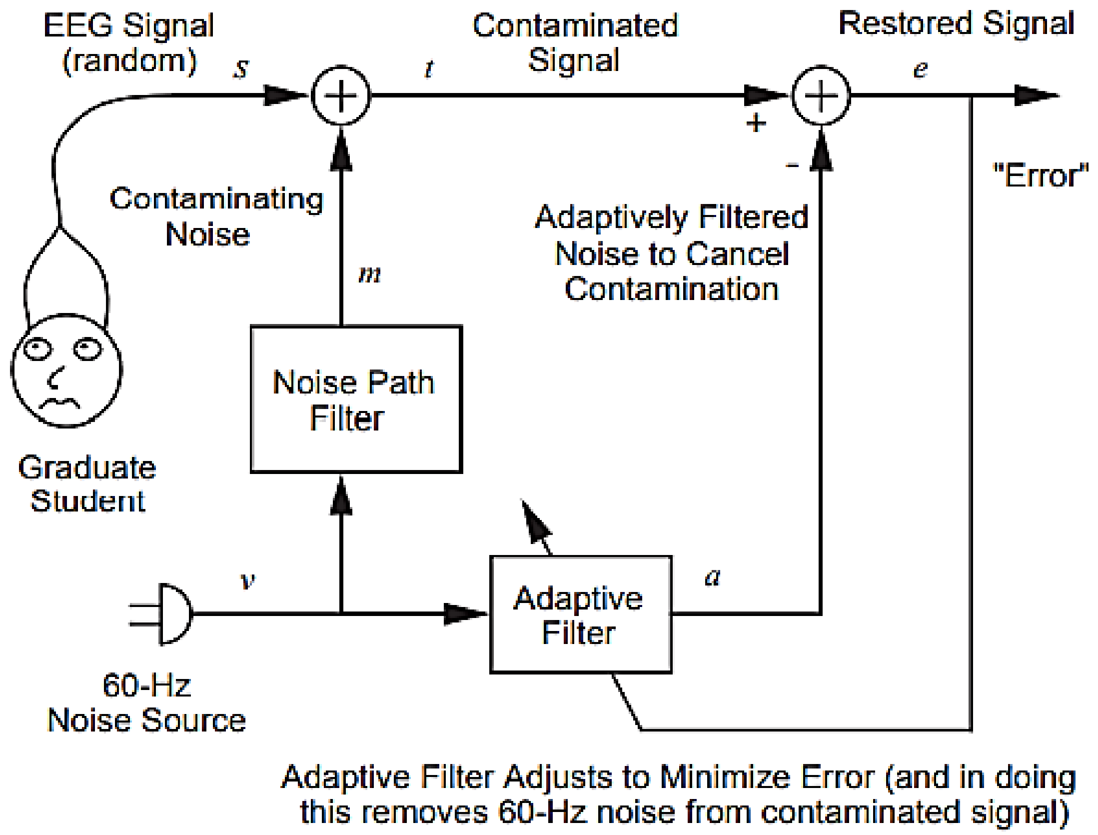
$y(k)$
$p_1(k) = y(k)$
D
$p_2(k) = y(k-1)$
D
D
$p_R(k) = y(k-R+1)$

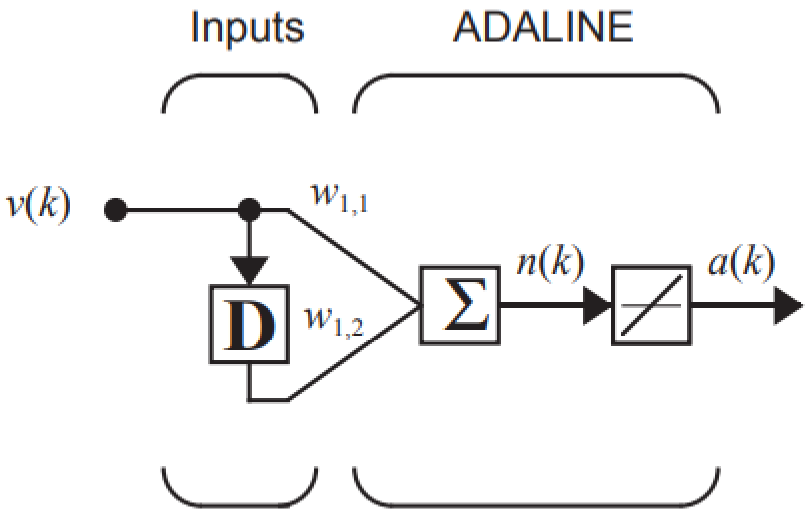
Бейімделетін сүзгі

$$a(k) = purelin(\mathbf{W}\mathbf{p} + b) = \sum_{i=1}^{R} w_{1,i}\, y(k-i+1) + b\,.$$



Егер сіз цифрлық сигналды өңдеумен таныс болсаңыз, 10.5-суреттегі желі жұмысын ақырғы импульстік жауап (FIR) сүзгісі ретінде танисыз. Цифрлық сигналды өңдеу саласын қарастыру бұл мәтіннің ауқымынан тыс, бірақ біз бұл бейімделгіш сүзгінің пайдалылығын қарапайым, бірақ практикалық мысал арқылы көрсете аламыз.





$$a(k) = w_{1,1}\, v(k) + w_{1,2}\, v(k-1)$$

$$\mathbf{R} = [\mathbf{z}\mathbf{z}^T] \text{ and } \mathbf{h} = E[t\mathbf{z}]\,.$$

$$\mathbf{z}(k) = \begin{bmatrix} v(k) \\ v(k-1) \end{bmatrix},$$

$$t(k) = s(k) + m(k)\,.$$

$$\mathbf{R} = \begin{bmatrix} E[v^2(k)] & E[v(k)v(k-1)] \\ E[v(k-1)v(k)] & E[v^2(k-1)] \end{bmatrix},$$

$$\mathbf{h} = \begin{bmatrix} E[(s(k)+m(k))v(k)] \\ E[(s(k)+m(k))v(k-1)] \end{bmatrix}.$$

$$v(k) = 1.2\,\sin\left(\frac{2\pi k}{3}\right),$$

$$m(k) = 0.12\,\sin\left(\frac{2\pi k}{3} + \frac{\pi}{2}\right).$$

$$E[v^2(k)] = (1.2)^2\frac{1}{3}\sum_{k=1}^{3}\left(\sin\left(\frac{2\pi k}{3}\right)\right)^2 = (1.2)^2 0.5 = 0.72,$$

$$E[v^2(k-1)] = E[v^2(k)] = 0.72,$$

$$E[v(k)v(k-1)] = \frac{1}{3}\sum_{k=1}^{3}\left(1.2\sin\frac{2\pi k}{3}\right)\left(1.2\sin\frac{2\pi(k-1)}{3}\right)$$

$$= (1.2)^2 0.5\cos\left(\frac{2\pi}{3}\right) = -0.36$$

$$\mathbf{R} = \begin{bmatrix} 0.72 & -0.36 \\ -0.36 & 0.72 \end{bmatrix}.$$

$$E[(s(k)+m(k))v(k)] = E[s(k)v(k)] + E[m(k)v(k)].$$

$$E[m(k)v(k)] = \frac{1}{3}\sum_{k=1}^{3}\left(0.12\ \sin\left(\frac{2\pi k}{3}+\frac{\pi}{2}\right)\right)\left(1.2\sin\frac{2\pi k}{3}\right) = 0$$

$$E[(s(k)+m(k))v(k-1)] = E[s(k)v(k-1)]$$
$$+E[m(k)v(k-1)]\ .$$

$$E[m(k)v(k-1)] = \frac{1}{3}\sum_{k=1}^{3}\left(0.12\ \sin\left(\frac{2\pi k}{3}+\frac{\pi}{2}\right)\right)\left(1.2\ \sin\frac{2\pi(k-1)}{3}\right)$$
$$= -0.0624\ .$$

$$\mathbf{h} = \begin{bmatrix} 0 \\ -0.0624 \end{bmatrix}.$$

$$\mathbf{x}^* = \mathbf{R}^{-1}\mathbf{h} = \begin{bmatrix} 0.72 & -0.36 \\ -0.36 & 0.72 \end{bmatrix}^{-1}\begin{bmatrix} 0 \\ -0.0624 \end{bmatrix} = \begin{bmatrix} -0.0578 \\ -0.1156 \end{bmatrix}.$$

$$F(\mathbf{x}) = c - 2\mathbf{x}^T\mathbf{h} + \mathbf{x}^T\mathbf{R}\mathbf{x}\ .$$

$$c = E[t^2(k)] = E[(s(k) + m(k))^2]$$

$$= E[s^2(k)] + 2E[s(k)m(k)] + E[m^2(k)] .$$

$$E[s^2(k)] = \frac{1}{0.4}\int_{-0.2}^{0.2} s^2 ds = \frac{1}{3(0.4)} s^3 \Big|_{-0.2}^{0.2} = 0.0133 .$$

$$E[m^2(k)] = \frac{1}{3}\sum_{k=1}^{3}\left\{0.12\ \sin\left(\frac{2\pi}{3} + \frac{\pi}{2}\right)\right\}^2 = 0.0072 ,$$

$$c = 0.0133 + 0.0072 = 0.0205 .$$

$$F(\mathbf{x}^*) = 0.0205 - 2(0.0072) + 0.0072 = 0.0133 .$$