Кері таралу

Дәріс 9-10

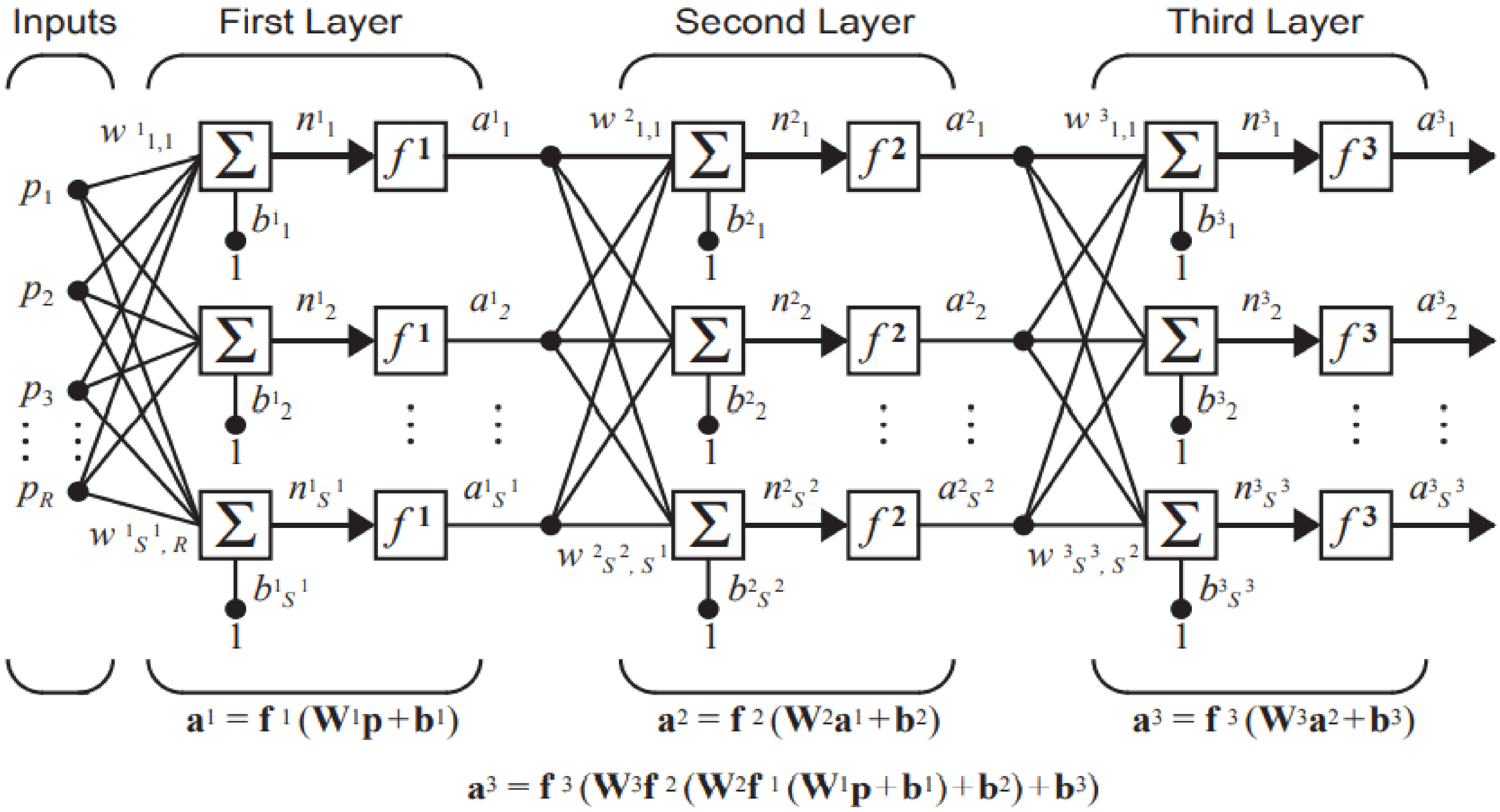
Кіріспе:

- Фрэнк Розенблаттың перцептронды үйрену ережесі және Бернард Уидроу мен Марсиан Хоффтың LMS алгоритмі бір қабатты перцептрон тәрізді желілерді оқытуға арналған. Алдыңғы дәрістерде талқылағанымыздай, бұл бірқабатты желілердің кемшілігі бар, олар тек сызықты түрде бөлінетін жіктеу мәселелерін шеше алады. Розенблат та, Уидроу да осы шектеулерді білетін және оларды жеңе алатын көпқабатты желілерді ұсынды, бірақ олар осы қуатты желілерді үйрету үшін алгоритмдерін жалпылай алмады.

Кіріспе:

- Көпқабатты желілік жұмыстарды үйрету алгоритмінің алғашқы сипаттамасы 1974 жылы Пол Вербостың тезисінде қамтылған болса керек [Werbo74]. Бұл жұмыс алгоритмді жалпы желілер контекстінде, нейрондық желілерді ерекше жағдай ретінде көрсетті және нейрондық желі қауымдастығында таралмады. 1980 жылдардың ортасында ғана кері таралу алгоритмі қайта ашылып, кеңінен жарияланды.

Көпқабатты перцептрондар



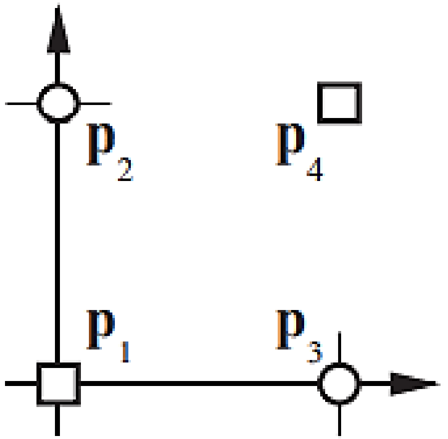
Көпқабатты перцептрондар

- Енді осы көпқабатты перцептрондық желілердің мүмкіндіктерін зерттейік. Алдымен үлгі классификациясы үшін көпқабатты желілерді пайдалануды қарастырамыз, содан кейін олардың функция жуықтауында қолданылуын талқылаймыз.

Үлгі классификациясы

- Үлгі классификациясы үшін көп қабатты перцептронның мүмкіндіктерін көрсету үшін классикалық эксклюзивті немесе (XOR) мәселесін қарастырыңыз. XOR элементы үшін кіріс/мақсат жұптары болып табылады

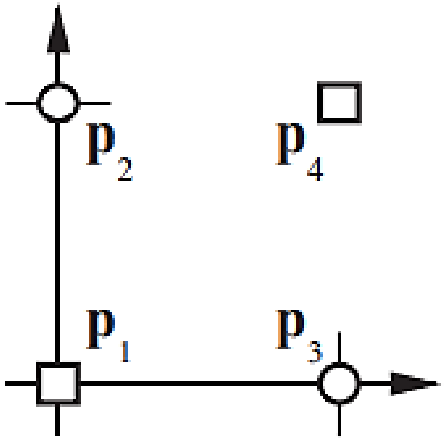
$$\left\{\mathbf{p}_1 = \begin{bmatrix}0\\0\end{bmatrix}, t_1 = 0\right\}\left\{\mathbf{p}_2 = \begin{bmatrix}0\\1\end{bmatrix}, t_2 = 1\right\}\left\{\mathbf{p}_3 = \begin{bmatrix}1\\0\end{bmatrix}, t_3 = 1\right\}\left\{\mathbf{p}_4 = \begin{bmatrix}1\\1\end{bmatrix}, t_4 = 0\right\}.$$

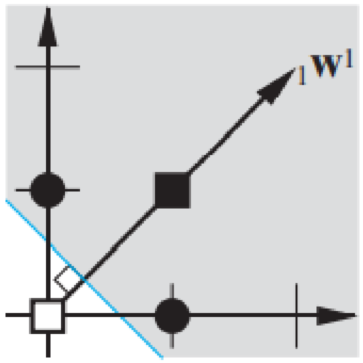


Үлгі классификациясы

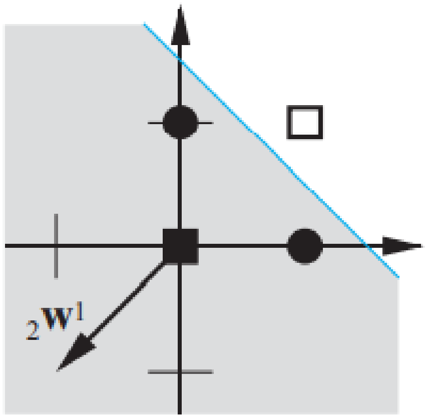
- Үлгі классификациясы үшін көп қабатты перцептронның мүмкіндіктерін көрсету үшін классикалық эксклюзивті немесе (XOR) мәселесін қарастырыңыз. XOR элементы үшін кіріс/мақсат жұптары болып табылады

$$\left\{\mathbf{p}_1 = \begin{bmatrix}0\\0\end{bmatrix}, t_1 = 0\right\}\left\{\mathbf{p}_2 = \begin{bmatrix}0\\1\end{bmatrix}, t_2 = 1\right\}\left\{\mathbf{p}_3 = \begin{bmatrix}1\\0\end{bmatrix}, t_3 = 1\right\}\left\{\mathbf{p}_4 = \begin{bmatrix}1\\1\end{bmatrix}, t_4 = 0\right\}.$$



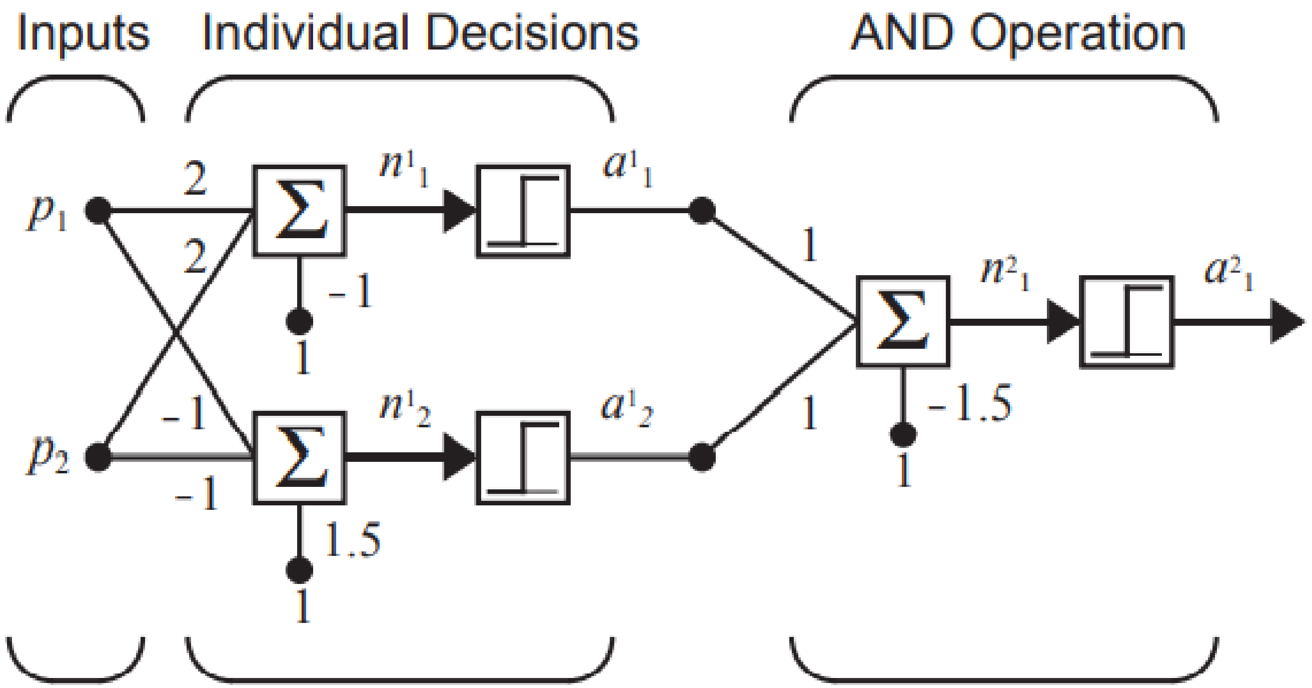


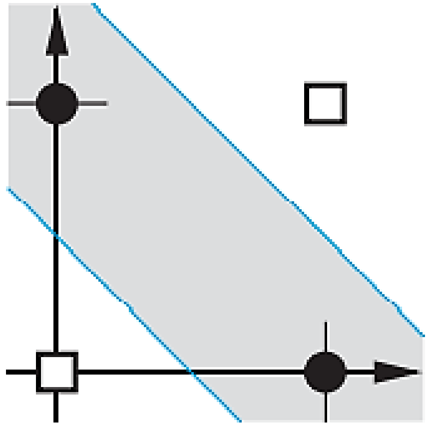
Layer 1/Neuron 1



Layer 1/Neuron 2

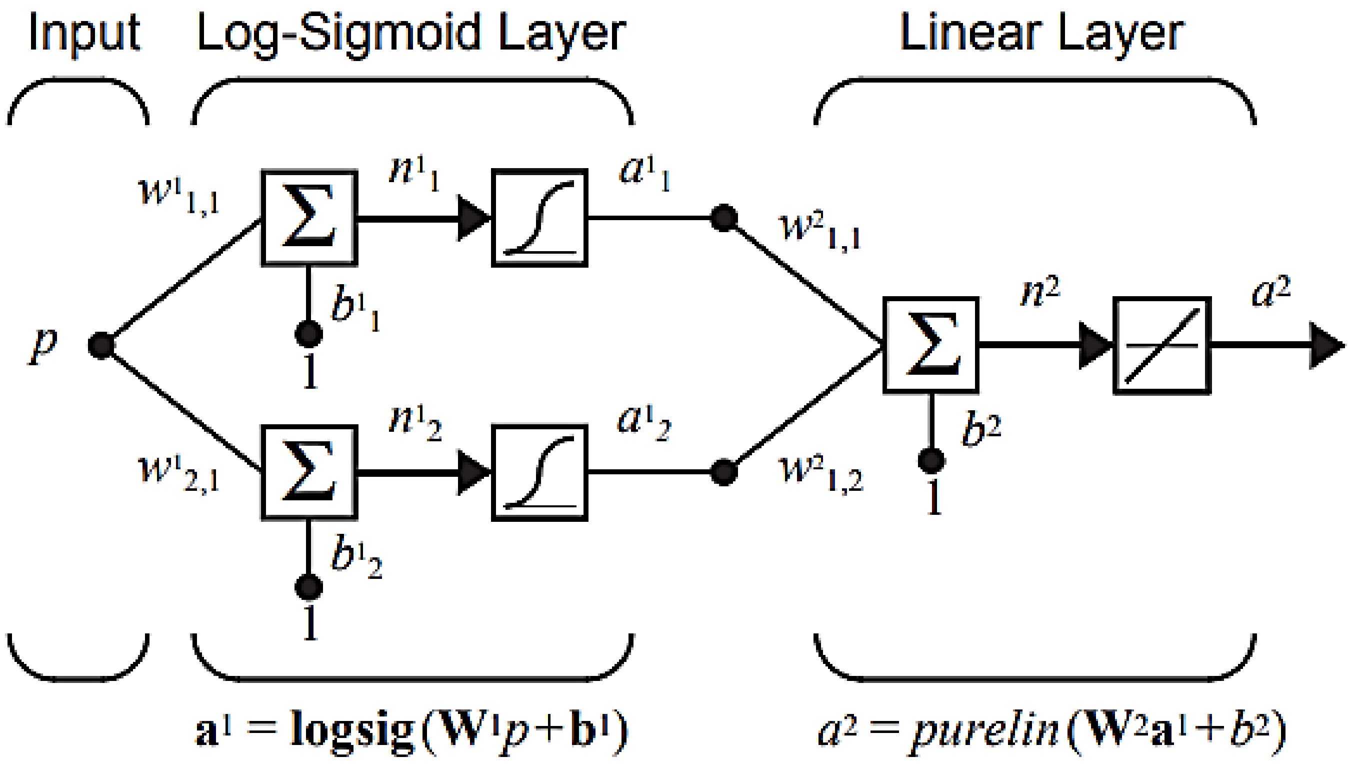
a(бит 1)	b(бит 2)	a(бит 1) ^ b(бит 2)
0	0	0
0	1	1
1	0	1
1	1	0





Функцияның жуықтауы

$$f^1(n) = \frac{1}{1+e^{-n}} \text{ and } f^2(n) = n .$$

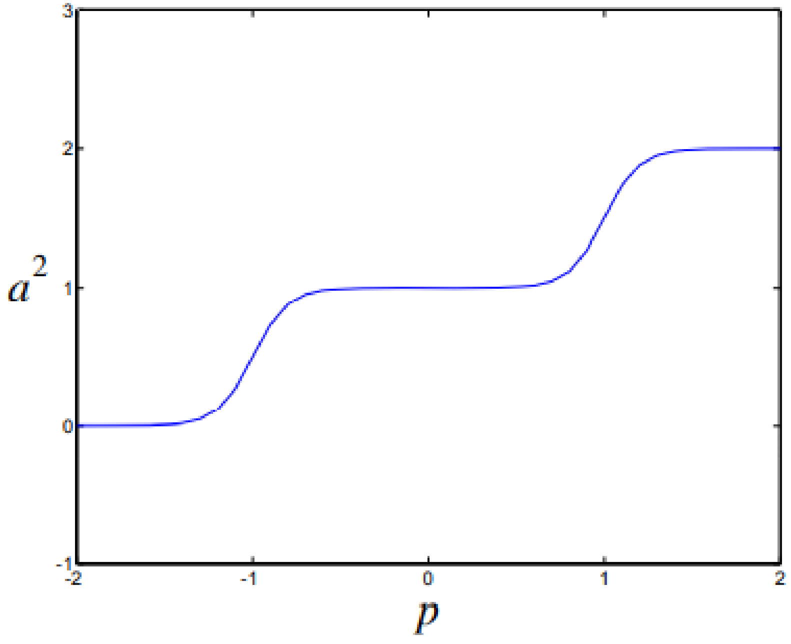


Функцияның жуықтауы

Бұл желі үшін салмақтар мен қиғаштықтардың номиналды мәндері делік

$$w^1_{1,1} = 10\,,\ w^1_{2,1} = 10\,,\ b^1_1 = -10\,,\ b^1_2 = 10\,,$$

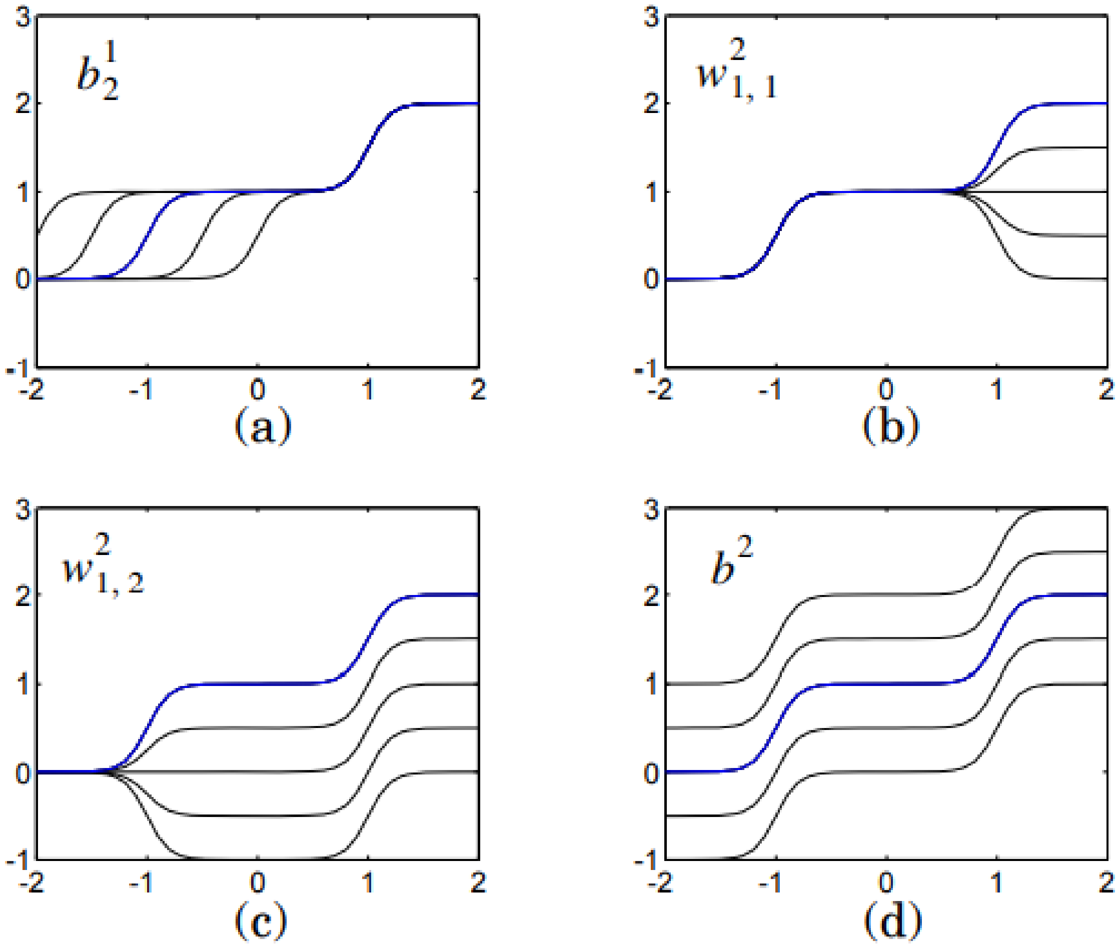
$$w^2_{1,1} = 1\,,\ w^2_{1,2} = 1\,,\ b^2 = 0\,.$$



$$n^1_2 = w^1_{2,1}p + b^1_2 = 0 \quad \Rightarrow \quad p = -\frac{b^1_2}{w^1_{2,1}} = -\frac{10}{10} = -1\,.$$

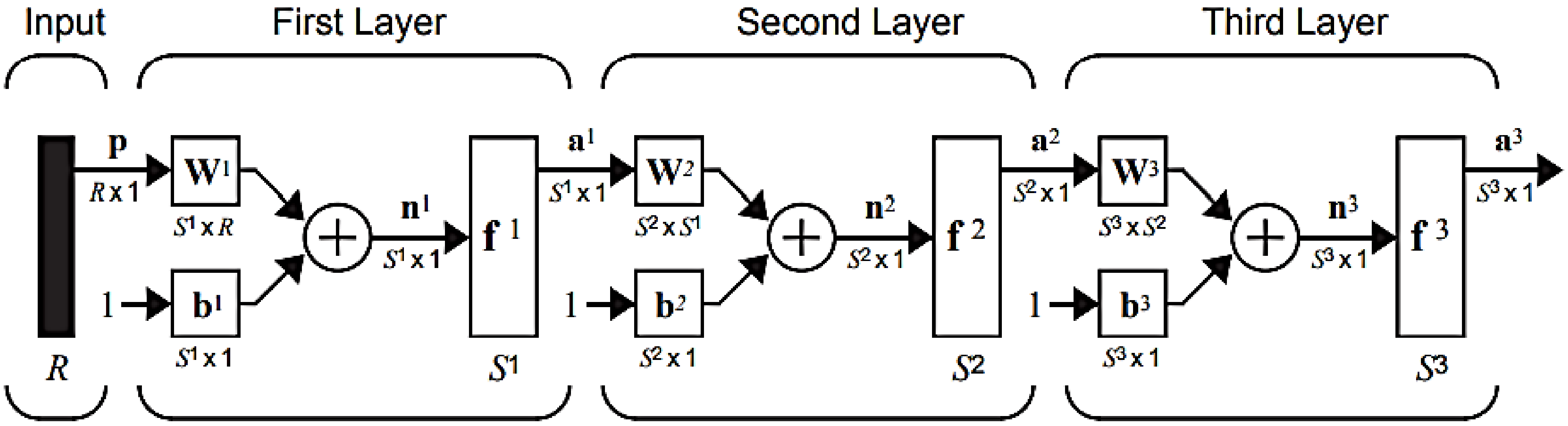
$$n^1_1 = w^1_{1,1}p + b^1_1 = 0 \quad \Rightarrow \quad p = -\frac{b^1_1}{w^1_{1,1}} = -\frac{-10}{10} = 1\,,$$

$$-1 \le w^2_{1,1} \le 1\ ,\ -1 \le w^2_{1,2} \le 1\ ,\ 0 \le b^1_2 \le 20\ ,\ -1 \le b^2 \le 1\ .$$



Енді біз үлгіні тану және функцияларды жақындату үшін көп қабатты перцептрондық желілердің күші туралы біраз түсінікке ие болғаннан кейін, келесі қадам осындай желілерді оқыту алгоритмін әзірлеу болып табылады.

Кері таралу алгоритмі



$$\mathbf{a}^1 = \mathbf{f}^1(\mathbf{W}^1\mathbf{p}+\mathbf{b}^1) \qquad \mathbf{a}^2 = \mathbf{f}^2(\mathbf{W}^2\mathbf{a}^1+\mathbf{b}^2) \qquad \mathbf{a}^3 = \mathbf{f}^3(\mathbf{W}^3\mathbf{a}^2+\mathbf{b}^3)$$

$$\mathbf{a}^3 = \mathbf{f}^3(\mathbf{W}^3\mathbf{f}^2(\mathbf{W}^2\mathbf{f}^1(\mathbf{W}^1\mathbf{p}+\mathbf{b}^1)+\mathbf{b}^2)+\mathbf{b}^3)$$

$$\mathbf{a}^{m+1} = \mathbf{f}^{m+1}(\mathbf{W}^{m+1}\mathbf{a}^m + \mathbf{b}^{m+1}) \text{ for } m = 0, 1, \ldots, M-1,$$

Өнімділік индексі

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

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

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

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

$$w_{i,j}^m(k+1) = w_{i,j}^m(k) - \alpha\frac{\partial\hat{F}}{\partial w_{i,j}^m},$$

$$b_i^m(k+1) = b_i^m(k) - \alpha\frac{\partial\hat{F}}{\partial b_i^m},$$

Тізбек ережесі

- Бірқабатты сызықтық желі (ADALINE) үшін бұл ішінара туындылар теңдеу арқылы ыңғайлы түрде есептеледі. Көп қабатты желі үшін қате жасырын қабаттардағы салмақтардың айқын функциясы емес, сондықтан бұл туындылар оңай есептелмейді. Қате жасырын қабаттардағы салмақтардың жанама функциясы болғандықтан, біз туындыларды есептеу үшін есептеудің тізбекті ережесін қолданамыз. Тізбек ережесін қарастыру үшін бізде тек айнымалының айқын функциясы болып табылатын функция бар делік.

- Біз үшінші айнымалыға қатысты туындысын алғымыз келеді. Сонда тізбек ережесі:

$$\frac{df(n(w))}{dw} = \frac{df(n)}{dn} \times \frac{dn(w)}{dw}.$$

$f(n) = e^n$ and $n = 2w$, so that $f(n(w)) = e^{2w}$,

$$\frac{df(n(w))}{dw} = \frac{df(n)}{dn} \times \frac{dn(w)}{dw} = (e^n)(2).$$

$$\frac{\partial \hat{F}}{\partial w_{i,j}^m} = \frac{\partial \hat{F}}{\partial n_i^m} \times \frac{\partial n_i^m}{\partial w_{i,j}^m},$$

$$\frac{\partial \hat{F}}{\partial b_i^m} = \frac{\partial \hat{F}}{\partial n_i^m} \times \frac{\partial n_i^m}{\partial b_i^m}.$$

- Осы теңдеулердің әрқайсысының екінші мүшесін оңай есептеуге болады, өйткені қабатқа таза кіріс сол қабаттағы салмақтар мен ауытқудың айқын функциясы болып табылады:

$$n_i^m = \sum_{j=1}^{S^{m-1}} w_{i,j}^m a_j^{m-1} + b_i^m .$$

$$\frac{\partial n_i^m}{\partial w_{i,j}^m} = a_j^{m-1} , \frac{\partial n_i^m}{\partial b_i^m} = 1 .$$

$$s_i^m \equiv \frac{\partial \hat{F}}{\partial n_i^m} ,$$

$$\frac{\partial \hat{F}}{\partial w_{i,j}^{m}} = s_i^m a_j^{m-1} ,$$

$$\frac{\partial \hat{F}}{\partial b_i^{m}} = s_i^m .$$

Енді біз шамамен ең тік түсу алгоритмін келесідей өрнектей аламыз

$$w_{i,j}^{m}(k+1) = w_{i,j}^{m}(k) - \alpha s_i^m a_j^{m-1} ,$$

$$b_i^{m}(k+1) = b_i^{m}(k) - \alpha s_i^m .$$

$$\mathbf{W}^{m}(k+1) = \mathbf{W}^{m}(k) - \alpha \mathbf{s}^{m}(\mathbf{a}^{m-1})^{T} ,$$

$$\mathbf{b}^{m}(k+1) = \mathbf{b}^{m}(k) - \alpha \mathbf{s}^{m} ,$$

$$\mathbf{s}^{m} \equiv \frac{\partial \hat{F}}{\partial \mathbf{n}^{m}} = \begin{bmatrix} \dfrac{\partial \hat{F}}{\partial n_1^{m}} \\ \dfrac{\partial \hat{F}}{\partial n_2^{m}} \\ \vdots \\ \dfrac{\partial \hat{F}}{\partial n_{S^m}^{m}} \end{bmatrix} .$$

Сезімталдықтарды кері тарату

- Енді бізге сезімталдықтарды есептеу қалды, бұл тізбек ережесін басқа қолдануды талап етеді. Дәл осы процесс бізге кері таралу терминін береді, өйткені ол қабаттағы сезімталдық қабаттағы сезімталдықтан есептелетін қайталану қатынасын сипаттайды.
- Сезімталдықтардың қайталану қатынасын алу үшін келесі якобиан матрицаны қолданамыз:

$$\frac{\partial \mathbf{n}^{m+1}}{\partial \mathbf{n}^{m}} \equiv \begin{bmatrix} \frac{\partial n_1^{m+1}}{\partial n_1^m} & \frac{\partial n_1^{m+1}}{\partial n_2^m} & \cdots & \frac{\partial n_1^{m+1}}{\partial n_{S^m}^m} \\ \frac{\partial n_2^{m+1}}{\partial n_1^m} & \frac{\partial n_2^{m+1}}{\partial n_2^m} & \cdots & \frac{\partial n_2^{m+1}}{\partial n_{S^m}^m} \\ \vdots & \vdots & & \vdots \\ \frac{\partial n_{S^{m+1}}^{m+1}}{\partial n_1^m} & \frac{\partial n_{S^{m+1}}^{m+1}}{\partial n_2^m} & \cdots & \frac{\partial n_{S^{m+1}}^{m+1}}{\partial n_{S^m}^m} \end{bmatrix}.$$

$$\frac{\partial n_i^{m+1}}{\partial n_j^m} = \frac{\partial\left(\sum_{l=1}^{S^m} w_{i,l}^{m+1} a_l^m + b_i^{m+1}\right)}{\partial n_j^m} = w_{i,j}^{m+1} \frac{\partial a_j^m}{\partial n_j^m}$$

$$= w_{i,j}^{m+1} \frac{\partial f^m(n_j^m)}{\partial n_j^m} = w_{i,j}^{m+1} \dot{f}^m(n_j^m) \,,$$

$$\dot{f}^m(n_j^m) = \frac{\partial f^m(n_j^m)}{\partial n_j^m} \,.$$

$$\frac{\partial \mathbf{n}^{m+1}}{\partial \mathbf{n}^m} = \mathbf{W}^{m+1} \dot{\mathbf{F}}^m(\mathbf{n}^m) \,,$$

$$\dot{\mathbf{F}}^m(\mathbf{n}^m) = \begin{bmatrix} \dot{f}^m(n_1^m) & 0 & \dots & 0 \\ 0 & \dot{f}^m(n_2^m) & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \dot{f}^m(n_{S^m}^m) \end{bmatrix} .$$

$$\mathbf{s}^m = \frac{\partial \hat{F}}{\partial \mathbf{n}^m} = \left(\frac{\partial \mathbf{n}^{m+1}}{\partial \mathbf{n}^m} \right)^T \frac{\partial \hat{F}}{\partial \mathbf{n}^{m+1}} = \dot{\mathbf{F}}^m(\mathbf{n}^m)(\mathbf{W}^{m+1})^T \frac{\partial \hat{F}}{\partial \mathbf{n}^{m+1}}$$

$$= \dot{\mathbf{F}}^m(\mathbf{n}^m)(\mathbf{W}^{m+1})^T \mathbf{s}^{m+1} .$$

$$\mathbf{s}^M \to \mathbf{s}^{M-1} \to \ldots \to \mathbf{s}^2 \to \mathbf{s}^1 .$$

$$s_i^M = \frac{\partial \hat{F}}{\partial n_i^M} = \frac{\partial (\mathbf{t}-\mathbf{a})^T(\mathbf{t}-\mathbf{a})}{\partial n_i^M} = \frac{\partial \sum_{j=1}^{S^M} (t_j - a_j)^2}{\partial n_i^M} = -2(t_i - a_i)\frac{\partial a_i}{\partial n_i^M} .$$

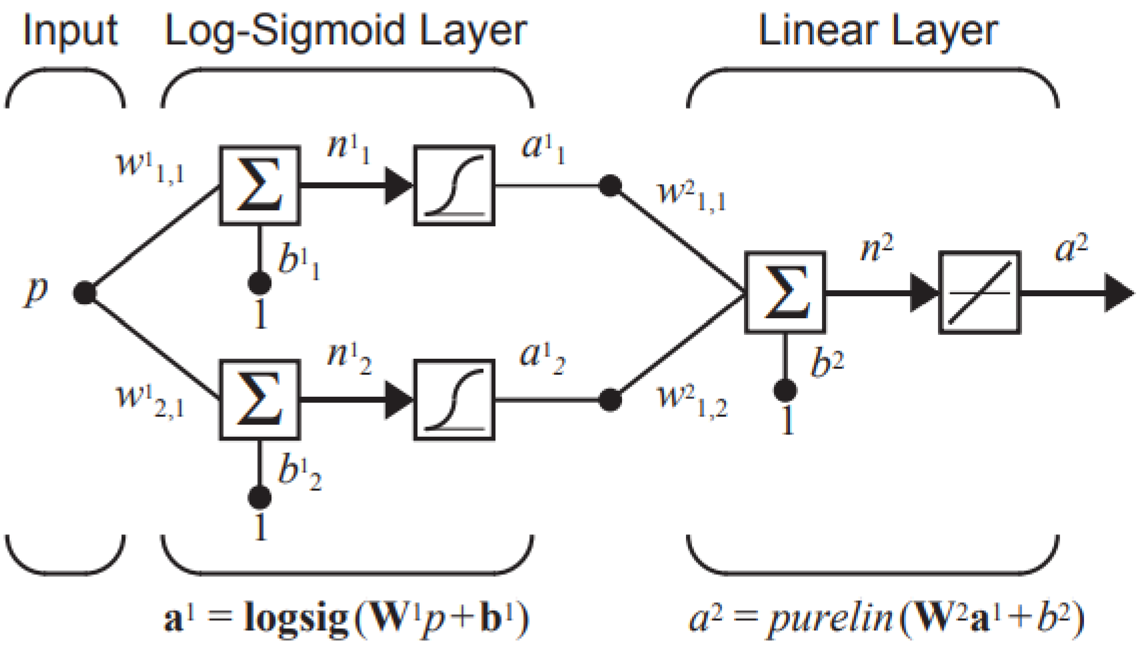
$$\frac{\partial a_i}{\partial n_i^M} = \frac{\partial a_i^M}{\partial n_i^M} = \frac{\partial f^M(n_i^M)}{\partial n_i^M} = \dot{f}^M(n_i^M) ,$$

$$s_i^M = -2(t_i - a_i)\dot{f}^M(n_i^M) .$$

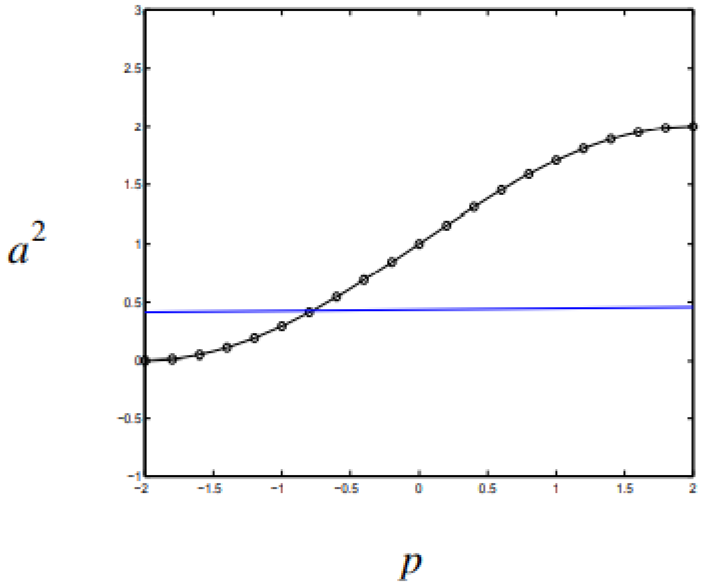
$$\mathbf{s}^M = -2\dot{\mathbf{F}}^M(\mathbf{n}^M)(\mathbf{t}-\mathbf{a}) .$$

Мысал,

$$g(p) = 1 + \sin\left(\frac{\pi}{4}p\right) \text{ for } -2 \le p \le 2 .$$



$$\mathbf{W}^1(0) = \begin{bmatrix} -0.27 \\ -0.41 \end{bmatrix},\ \mathbf{b}^1(0) = \begin{bmatrix} -0.48 \\ -0.13 \end{bmatrix},\ \mathbf{W}^2(0) = \begin{bmatrix} 0.09 & -0.17 \end{bmatrix},\ \mathbf{b}^2(0) = \begin{bmatrix} 0.48 \end{bmatrix}.$$



$$\mathbf{a}^1 = \mathbf{f}^1(\mathbf{W}^1\mathbf{a}^0 + \mathbf{b}^1) = \mathbf{logsig}\left(\begin{bmatrix}-0.27\\-0.41\end{bmatrix}\begin{bmatrix}1\end{bmatrix} + \begin{bmatrix}-0.48\\-0.13\end{bmatrix}\right) = \mathbf{logsig}\left(\begin{bmatrix}-0.75\\-0.54\end{bmatrix}\right)$$

$$= \begin{bmatrix}\dfrac{1}{1+e^{0.75}}\\[2ex]\dfrac{1}{1+e^{0.54}}\end{bmatrix} = \begin{bmatrix}0.321\\0.368\end{bmatrix}.$$

$$a^2 = f^2(\mathbf{W}^2\mathbf{a}^1 + \mathbf{b}^2) = purelin\left(\begin{bmatrix}0.09 & -0.17\end{bmatrix}\begin{bmatrix}0.321\\0.368\end{bmatrix} + \begin{bmatrix}0.48\end{bmatrix}\right) = \begin{bmatrix}0.446\end{bmatrix}.$$

$$e = t - a = \left\{1 + \sin\left(\frac{\pi}{4}p\right)\right\} - a^2 = \left\{1 + \sin\left(\frac{\pi}{4}1\right)\right\} - 0.446 = 1.261.$$

$$\dot{f}^1(n) = \frac{d}{dn}\left(\frac{1}{1+e^{-n}}\right) = \frac{e^{-n}}{(1+e^{-n})^2} = \left(1 - \frac{1}{1+e^{-n}}\right)\left(\frac{1}{1+e^{-n}}\right) = (1-a^1)(a^1).$$

$$\dot{f}^2(n) = \frac{d}{dn}(n) = 1.$$

$$\mathbf{s}^2 = -2\dot{\mathbf{F}}^2(\mathbf{n}^2)(\mathbf{t}-\mathbf{a}) = -2\left[\dot{f}^2(n^2)\right](1.261) = -2\left[1\right](1.261) = -2.522 .$$

$$\mathbf{s}^1 = \dot{\mathbf{F}}^1(\mathbf{n}^1)(\mathbf{W}^2)^T\mathbf{s}^2 = \begin{bmatrix} (1-a_1^1)(a_1^1) & 0 \\ 0 & (1-a_2^1)(a_2^1) \end{bmatrix} \begin{bmatrix} 0.09 \\ -0.17 \end{bmatrix} \begin{bmatrix} -2.522 \end{bmatrix}$$

$$= \begin{bmatrix} (1-0.321)(0.321) & 0 \\ 0 & (1-0.368)(0.368) \end{bmatrix} \begin{bmatrix} 0.09 \\ -0.17 \end{bmatrix} \begin{bmatrix} -2.522 \end{bmatrix}$$

$$= \begin{bmatrix} 0.218 & 0 \\ 0 & 0.233 \end{bmatrix} \begin{bmatrix} -0.227 \\ 0.429 \end{bmatrix} = \begin{bmatrix} -0.0495 \\ 0.0997 \end{bmatrix} .$$

$$\mathbf{W}^2(1) = \mathbf{W}^2(0) - \alpha\mathbf{s}^2(\mathbf{a}^1)^T = \begin{bmatrix} 0.09 & -0.17 \end{bmatrix} - 0.1\begin{bmatrix} -2.522 \end{bmatrix}\begin{bmatrix} 0.321 & 0.368 \end{bmatrix}$$

$$= \begin{bmatrix} 0.171 & -0.0772 \end{bmatrix},$$

$$\mathbf{b}^2(1) = \mathbf{b}^2(0) - \alpha\mathbf{s}^2 = \begin{bmatrix} 0.48 \end{bmatrix} - 0.1\begin{bmatrix} -2.522 \end{bmatrix} = \begin{bmatrix} 0.732 \end{bmatrix},$$

$$\mathbf{W}^1(1) = \mathbf{W}^1(0) - \alpha\mathbf{s}^1(\mathbf{a}^0)^T = \begin{bmatrix} -0.27 \\ -0.41 \end{bmatrix} - 0.1\begin{bmatrix} -0.0495 \\ 0.0997 \end{bmatrix}\begin{bmatrix} 1 \end{bmatrix} = \begin{bmatrix} -0.265 \\ -0.420 \end{bmatrix},$$

$$\mathbf{b}^1(1) = \mathbf{b}^1(0) - \alpha\mathbf{s}^1 = \begin{bmatrix} -0.48 \\ -0.13 \end{bmatrix} - 0.1\begin{bmatrix} -0.0495 \\ 0.0997 \end{bmatrix} = \begin{bmatrix} -0.475 \\ -0.140 \end{bmatrix}.$$