2021. 02. 11.

Perceptron kapacitás:

$$c(p|N) = L(p|N) / 2^p$$

$$c(p|N) = \begin{cases} 1 & | \quad p \leq N \\ \frac{2}{2^p} \cdot \sum_{t=0}^{N-1} \binom{p-1}{t} & | \quad p > N \end{cases}$$

Ha $p \leq N$

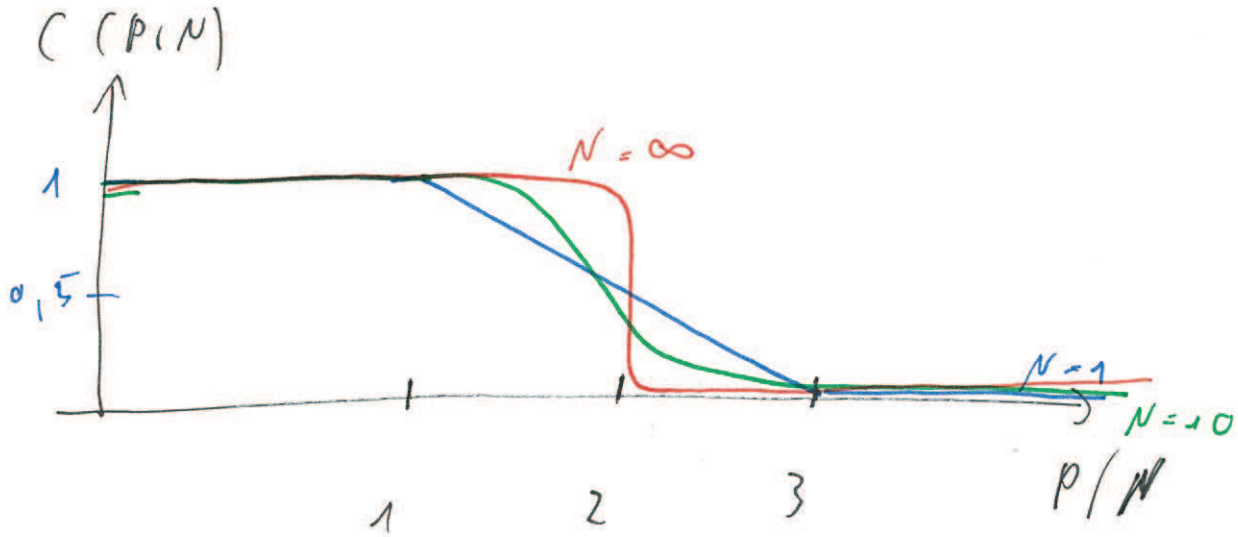
$$\underline{\underline{X}} = \begin{bmatrix} — & x_1^T & — & 1 \\ — & x_2^T & — & 1 \\ & \vdots & & \\ — & x_p^T & — & 1 \end{bmatrix} \qquad \underline{d} = \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_p \end{bmatrix}$$

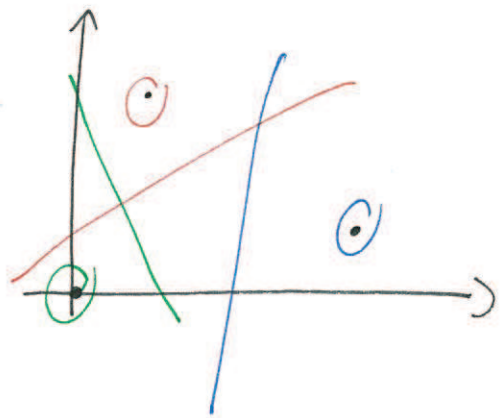
$$\mathrm{sign}(\underline{\underline{X}} \cdot \underline{w}) = \underline{d}$$

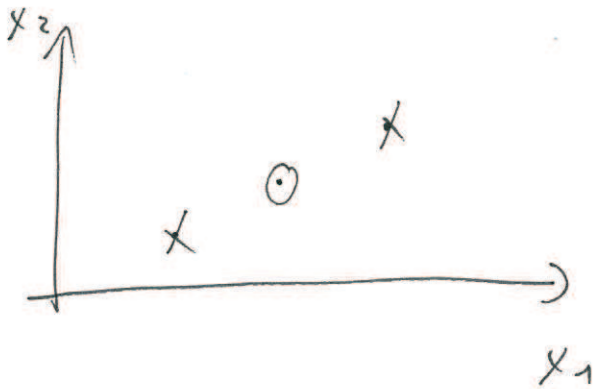
$$[X_k^T | 1] = \alpha_1 \cdot [X_1^T, 1] + \alpha_2 [X_2^T, 1] + \dots + \alpha_{k-1} \cdot [X_{k-1}^T | 1]$$

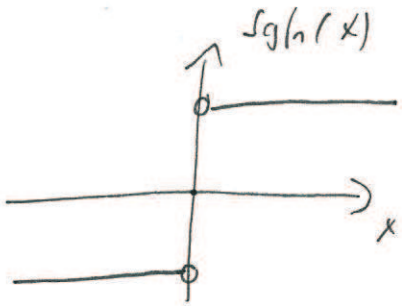
$\alpha_i > 0 \Rightarrow d_i = 1$ eselt. $d_k = -1$

$\alpha_i < 0 \Rightarrow d_i = -1$



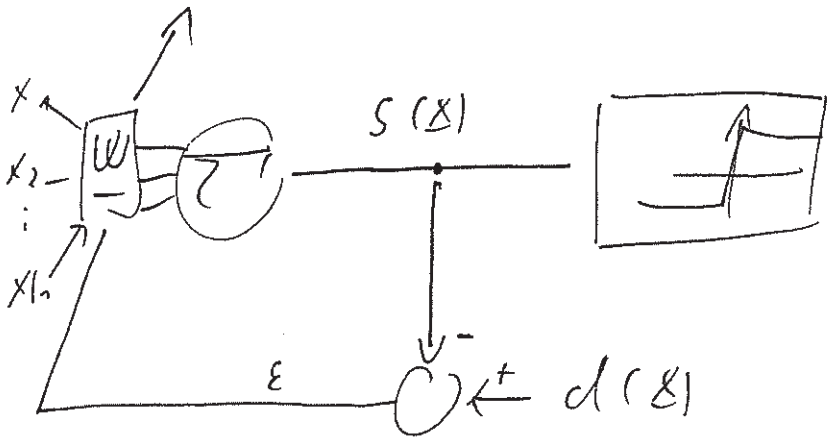






Adaptive Linear Neuron (Adaline):

$$y(\underline{x}) = \operatorname{sgn}(\underline{w}^T \underline{x})$$



$$C(\underline{w}) = \mathbb{E}_{(\underline{x},d)\sim\xi}\left\{(d - \underline{w}^T\underline{x})^2\right\} = \mathbb{E}\{d^2\} + \mathbb{E}\{\underline{w}^T \cdot \underline{x} \cdot \underline{x}^T \cdot \underline{w}\} - 2\,\mathbb{E}\{\underline{w}^T \cdot \underline{x} \cdot d\}$$

$$\underline{w}^* = \arg\min_{\underline{w}} \{C(\underline{w})\}$$

$$\underline{\underline{R}}_{x\hat{x}} = (1/N) \cdot \underline{\underline{X}}^T \cdot \underline{\underline{x}} \qquad \hat{\underline{p}} = (1/N)\, \underline{\underline{X}}^T \underline{d}$$

$$C(\underline{w}) = \lambda + \underline{w}^T \cdot \underbrace{\mathbb{E}\{\underline{x} \cdot \underline{x}^T\}}_{\underline{\underline{R}}_{xx}} \underline{w} - 2\, \underline{w}^T \cdot \underbrace{\mathbb{E}\{\underline{x} \cdot d\}}_{\underline{p}}$$

$\underline{\underline{R}}_{xx}$ autokorrelációs mátrix.

$\underline{p}$: keresztkorrelációs vektor

$$\underline{\underline{R}}_{xx} \succeq \underline{\underline{0}} \iff \forall\, \underline{v} \in \mathbb{R}^N \quad \underline{v}^T \underline{\underline{R}}_{xx} \underline{v} \geq 0$$

$$\underline{v}^T \underline{\underline{R}}_{xx} \underline{v} = \underline{v}^T \cdot \mathbb{E}\{\underline{x} \cdot \underline{x}^T\} \cdot \underline{v} = \mathbb{E}\{\underline{v}^T \underline{x} \cdot \underline{x}^T \cdot \underline{v}\} =$$

$$= \mathbb{E}\{(\underline{v}^T \cdot \underline{x})^2\} \geq 0 \Rightarrow C \text{ konvex}$$

$$\underline{\nabla C(\underline{w}^*) = \underline{0}}$$

$$\nabla C(\underline{w}) = 2\, \underline{\underline{R}}_{xx} \cdot \underline{w} - 2 \cdot \underline{p} = \underline{0}$$

$$\boxed{\underline{w}^* = \underline{\underline{R}}_{xx}^{-1}\, \underline{p}}$$

Wiener-Hopf

$$C(\underline{w}) = \alpha + \underline{w}^T \underline{\underline{R}}_{xx} \underline{w} - 2\underline{w}^T \underline{p}$$

$$C^{(2)}_{\underline{w}^*}(\underline{w}) = C(\underline{w}^*) + (\underline{w} - \underline{w}^*)^T \cdot \nabla C(\underline{w}^*) + (1/2)(\underline{w} - \underline{w}^*)^T \cdot \underline{\underline{\frac{\partial^2 C(\underline{w}^*)}{\partial w^2}}} \cdot (\underline{w} - \underline{w}^*)$$

$$\underline{\underline{\frac{\partial^2 C(\underline{w}^*)}{\partial w^2}}}(i,j) = \frac{\partial^2 C(\underline{w}^*)}{\partial w_i \, \partial w_j}$$

$$C(\underline{w}) = C(\underline{w}^*) + (\underline{w} - \underline{w}^*)^T \cdot (2\underline{\underline{R}}_{xx} \underline{w}^* - 2\underline{p}) + 1/2 (\underline{w} - \underline{w}^*)^T \cdot 2\underline{\underline{R}}_{xx} \cdot (\underline{w} - \underline{w}^*)$$

$$\underbrace{\underline{w}^* = \underline{\underline{R}}_{xx}^{-1} \underline{p}}_{0}$$

$$C(\underline{w}) = C(\underline{w}^*) + (\underline{w} - \underline{w}^*)^T \underline{\underline{R}}_{xx} (\underline{w} - \underline{w}^*)$$

$$\underline{\underline{R}}_{xx} = \mathbb{E}\{\underline{x} \cdot \underline{x}^T\} \qquad \underline{\underline{R}}_{xx}(i,j) = \mathbb{E}\{x_i \, x_j\} = \mathbb{E}\{x_j \cdot x_i\} = \underline{\underline{R}}_{xx}(j,i)$$

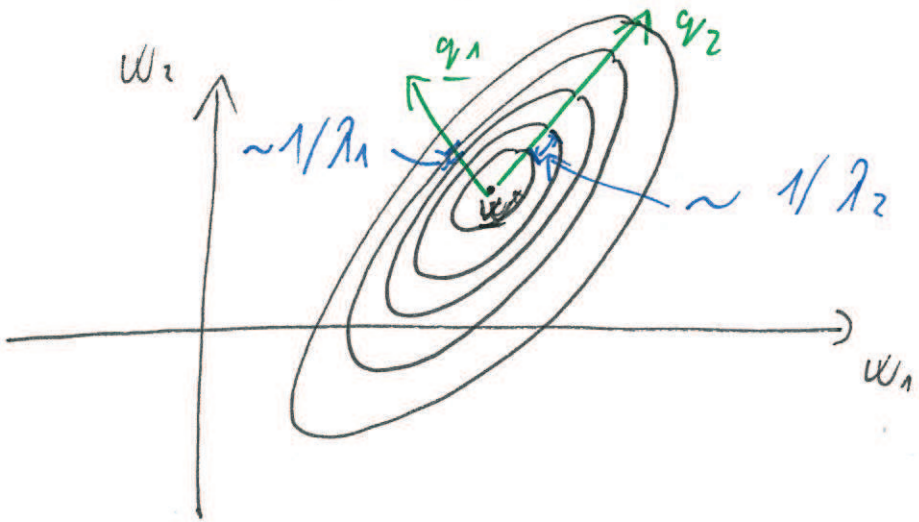
$$\underline{\underline{R}}_{xx} \succeq \underline{\underline{0}}$$

$$\underline{\underline{R_{xx}}} = \underline{\underline{Q}} \cdot \underline{\underline{\Lambda}} \cdot \underline{\underline{Q}}^T \qquad \underline{\underline{Q}} = \begin{bmatrix} | & | & & | \\ q_1 & q_2 & \cdots & q_n \\ | & | & & | \end{bmatrix} \qquad \underline{\underline{\Lambda}} = \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \ddots \; \lambda_n \end{bmatrix}$$

$$\underline{\underline{Q}}^{-1} = \underline{\underline{Q}}^T \qquad \forall i \;\; \lambda_i \geq 0$$

$$\left(\underbrace{\underline{v}^T \underline{\underline{Q}}}_{\underline{v}'^T} \cdot \underline{\underline{\Lambda}} \cdot \underbrace{\underline{\underline{Q}}^T \underline{v}}_{\underline{v}'} \geq 0 \right)$$

$$C(\underline{w}) = C(\underline{w}^*) + (\underline{w} - \underline{w}^*)^T \cdot \underline{\underline{Q}} \cdot \underline{\underline{\Lambda}} \cdot \underline{\underline{Q}}^T \cdot (\underline{w} - \underline{w}^*)$$



Gradient Descent alapú optimalizáció:

$$C(\underline{w}) = J + \underline{w}^T \underline{\underline{R}}_{xx} \underline{w} - 2\underline{w}^T \cdot \underline{p}$$

$$\underline{w}(k+1) = \underline{w}(k) - \mu \cdot \nabla C(\underline{w}(k)) \qquad \mu \in \mathbb{R}_{++}$$

$$\underline{w}(k+1) - \underline{w}^* = \underline{w}(k) - \underline{w}^* - \mu(\nabla C(\underline{w}(k)))$$

$$\underline{w}(k+1) - \underline{w}^* = \underline{w}(k) - \underline{w}^* - \mu(2\underline{\underline{R}}_{xx}\underline{w}(k) - 2\underbrace{\underline{p}}_{\underline{\underline{R}}_{xx}\cdot\underline{w}^*}) \qquad (\underline{w}^* = \underline{\underline{R}}_{xx}^{-1}\underline{p})$$

$$\underline{w}(k+1) - \underline{w}^* = \underline{w}(k) - \underline{w}^* - 2\mu \cdot \underline{\underline{R}}_{xx}(\underline{w}(k) - \underline{w}^*)$$

$$\underline{w}(k+1) - \underline{w}^* = (\underline{\underline{I}} - 2\mu \cdot \underline{\underline{R}}_{xx})(\underline{w}(k) - \underline{w}^*)$$

$$\forall\ \underline{w}(0) \qquad \lim_{k\to\infty} \underline{w}(k+1) - \underline{w}^* = \underline{0}$$

$$\underline{w}(k+1) - \underline{w}^* = (\underline{\underline{I}} - 2\mu\underline{\underline{R}}_{xx})^{k+1}(\underline{w}(0) - \underline{w}^*)$$

$$\underline{\underline{R}}_{xx} = \underline{\underline{Q}} \cdot \underline{\underline{\Lambda}} \cdot \underline{\underline{Q}}^T \qquad \text{da} \quad \underline{\underline{Q}} \cdot \underline{\underline{Q}}^T = \underline{\underline{I}}$$

$$\underline{w}(k+1) - \underline{w}^* = \left(\underline{\underline{Q}}\,\underline{\underline{Q}}^T - 2\mu \cdot \underline{\underline{Q}} \cdot \underline{\underline{\Lambda}} \cdot \underline{\underline{Q}}^T \right)^{k+1} \cdot \left(\underline{w}(0) - \underline{w}^* \right)$$

$$= \left(\underline{\underline{Q}} \cdot \underbrace{\left(\underline{\underline{I}} - 2\mu \cdot \underline{\underline{\Lambda}} \right)} \underline{\underline{Q}}^T \right)^{k+1} \cdot \left(\underline{w}(0) - \underline{w}^* \right)$$

$$\begin{bmatrix} 1-2\mu\lambda_1 & & 0 \\ & 1-2\mu\lambda_2 & \\ 0 & & \ddots \\ & & & 1-2\mu\lambda_n \end{bmatrix}$$

$$\underline{w}(k+1) - \underline{w}^* = \underline{\underline{Q}} \left(\underline{\underline{I}} - 2\mu \underline{\underline{\Lambda}} \right)^{k+1} \cdot \underline{\underline{Q}}^T \cdot \left(\underline{w}(0) - \underline{w}^* \right)$$

$$\begin{bmatrix} (1-2\mu\lambda_1)^{k+1} & & 0 \\ & (1-2\mu\lambda_2)^{k+1} & \\ 0 & & \ddots \\ & & & (1-2\mu\lambda_n)^{k+1} \end{bmatrix}$$

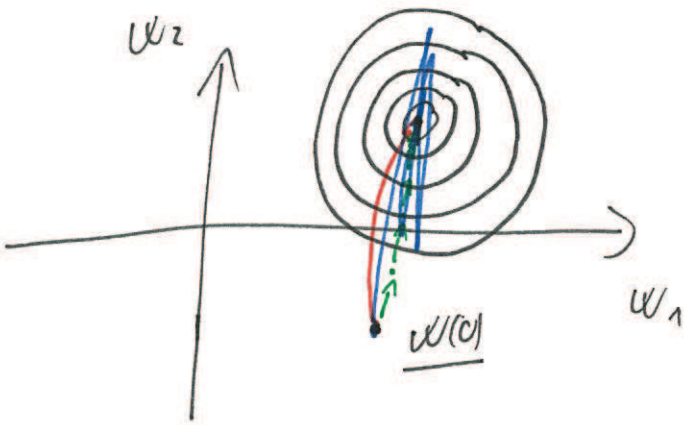
$$\forall i \quad |1 - 2\mu \cdot \lambda_i| < 1 \qquad 1 - 2\mu\lambda_i > -1$$

$$2 > 2\mu\lambda_i$$

$$\boxed{1/\lambda_{max} > \mu > 0} \quad \Longleftrightarrow \quad (\mu >) \quad \mu < 1/\lambda_i \quad \forall i$$

$$\underline{w}(k) - \underline{w}^* = \underline{\underline{Q}}\,(\underline{\underline{\Gamma}} - 2\mu\cdot\underline{\underline{\Lambda}})\,\underline{\underline{Q}}^T\,(\underline{w}(k-1) - \underline{w}^*)$$

T. f. h. $\lambda_1 = \lambda_2$



$$\boxed{0 < \mu < 1/(2\lambda_1)}$$

$(1/2\lambda_1) = \mu$

~~$1/(2\lambda_1)$~~ $1/2\lambda_1 < \mu < 1/\lambda_1$