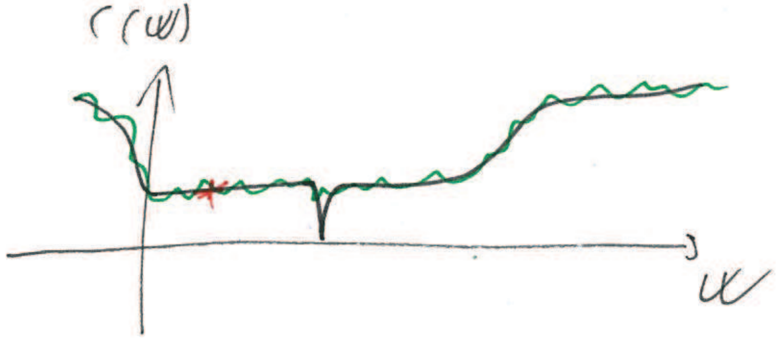
– Sztochasztikus gradiens módszerek:

- batch méret kérdése
- epoch



$$f(\underline{x}, \underline{w}) = y(\underline{x})$$

– Korai leállásos tanítás:

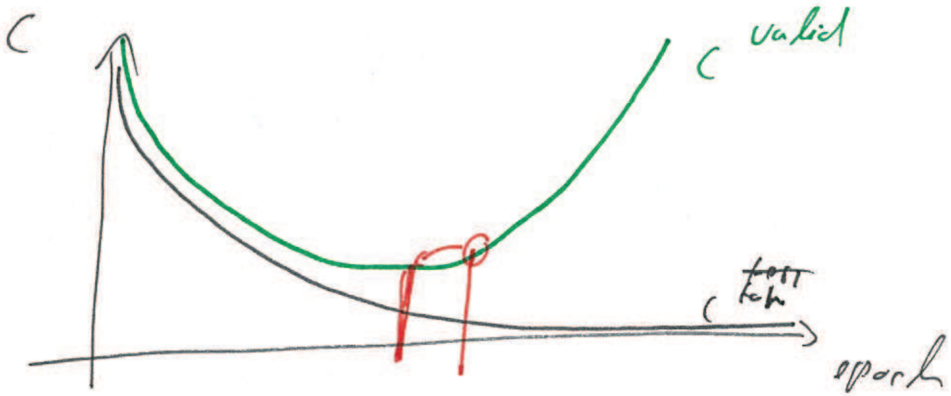
$$X = \{(x_i d)_i\}_i$$

$$X^{tan} \cup X^{teszt} \cup X^{valid} = X$$

$$\forall\, i \neq j \quad X^i \cap X^j = \cancel{\{0\}}\ \emptyset$$

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

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



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

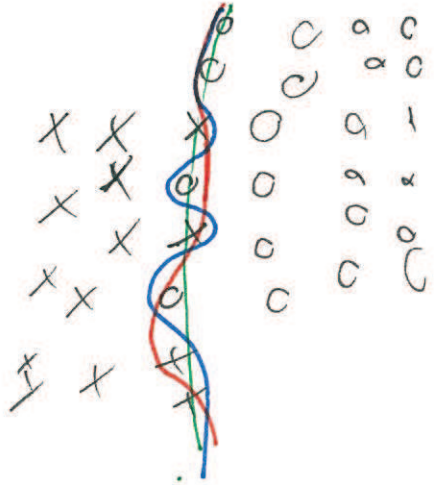
$$\underline{\underline{R}}_{xx} = \underline{\underline{Q}} \cdot \underline{\underline{\Lambda}} \cdot \underline{\underline{Q}}^T \qquad \underline{\underline{\Lambda}} = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix}$$

$$\underline{\underline{R}}_{xx}^{-1} = \underline{\underline{Q}} \cdot \underline{\underline{\Lambda}}^{-1} \cdot \underline{\underline{Q}}^T \qquad \underline{\underline{\Lambda}}^{-1} = \begin{bmatrix} 1/\lambda_1 & & \\ & \ddots & \\ & & 1/\lambda_n \end{bmatrix}$$

$$\underline{w} = \underline{\underline{R}}_{xx}^{-1} \cdot (\underline{p} + \underline{\varepsilon}) \qquad \forall i \quad |\underline{q}_i^T \cdot \underline{\varepsilon}| = 1$$

$$\underline{w} = \underline{w}^* + \underline{\underline{Q}} \cdot \begin{bmatrix} 1/\lambda_1 & & \\ & \ddots & \\ & & 1/\lambda_n \end{bmatrix} \cdot \begin{bmatrix} \underline{q}_1^T \underline{\varepsilon} \\ \underline{q}_2^T \underline{\varepsilon} \\ \vdots \\ \underline{q}_n^T \underline{\varepsilon} \end{bmatrix}$$

$$\|\underline{w} - \underline{w}^*\|_2^2 = \begin{bmatrix} \underline{q}_1^T \underline{\varepsilon} \\ \underline{q}_2^T \underline{\varepsilon} \\ \vdots \\ \underline{q}_n^T \underline{\varepsilon} \end{bmatrix}^T \cdot \overbrace{\underline{\underline{\Lambda}}^{-1} \cdot \underbrace{\underline{\underline{Q}}^T \cdot \underline{\underline{Q}}}_{\underline{\underline{I}}} \cdot \underline{\underline{\Lambda}}^{-1}}^{\begin{bmatrix} 1/\lambda_1^2 & & & \\ & 1/\lambda_2^2 & & \\ & & \ddots & \\ & & & 1/\lambda_n^2 \end{bmatrix}} \cdot \begin{bmatrix} \underline{q}_1 \underline{\varepsilon} \\ \vdots \\ \underline{q}_n^T \underline{\varepsilon} \end{bmatrix} = \sum_i (\underline{q}_i^T \underline{\varepsilon})^2 / \lambda_i^2$$



Adatok augmentációja

Bázis függvényes hálók:

- Bemenet transzformálása: $\varphi: \mathbb{R}^N \to \mathbb{R}^M \quad M > N$

- $y(\underline{x}) = \underline{\varphi(x)}^T \cdot \underline{w}$

$$\underline{\underline{\Phi}} = \underbrace{\begin{bmatrix} -\underline{\varphi(x_1)}^T- \\ -\underline{\varphi(x_2)}^T- \\ \vdots \\ -\underline{\varphi(x_n)}^T- \end{bmatrix}}_{M}$$

$$\underline{w}^* = \arg\min_{\underline{w}} \left\{ \| \underline{d} - \underline{\underline{\Phi}}\, \underline{w} \|_2^2 \right\}$$

$M \geq n$

$$C(\underline{w}) = \underline{d}^T \underline{d} - 2\,\underline{w}^T \underline{\underline{\Phi}}^T \underline{d} + \underline{w}^T \underline{\underline{\Phi}}^T \underline{\underline{\Phi}}\,\underline{w}$$

$$\frac{\partial C(\underline{w})}{\partial \underline{w}} = \underline{0} \qquad -2\,\underline{\underline{\Phi}}^T \underline{d} + 2\,\underline{\underline{\Phi}}^T \underline{\underline{\Phi}}\,\underline{w} = \underline{0}$$

$$\exists\, \underline{\underline{\Phi}}^T \underline{\underline{\Phi}}^{-1} \qquad \underline{w}^* = \underbrace{\left(\underline{\underline{\Phi}}^T \underline{\underline{\Phi}}\right)^{-1} \underline{\underline{\Phi}}^T}\,\underline{d} = \underline{\underline{\Phi}}^+ \underline{d}$$

Moore – Penrose pseudo inverz

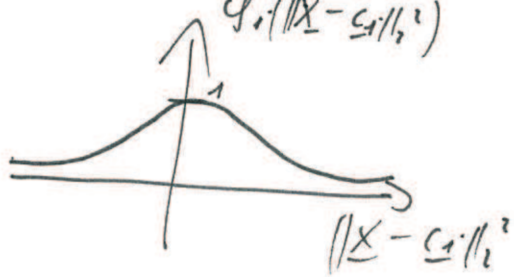
$$\nexists\, \underline{\underline{\Phi}}^T \underline{\underline{\Phi}}^{-1} : \qquad \underline{w}^* = \left(\underline{\underline{\Phi}}^T \underline{\underline{\Phi}} + \lambda \cdot \underline{\underline{I}}\right)^{-1} \underline{d}$$

$$\operatorname{rang}\left(\underline{\underline{\Phi}}^T \underline{\underline{\Phi}}\right) \leq n$$

$$M := n$$

<u>Radiális bázisfgv.-es hálók</u>

$$\varphi_i(\underline{x}) = \exp\left(-\left(1/(2\sigma^2)\right) \cdot \|\underline{x} - \cancel{kk}\ \underline{c}_i\|_2^2\right)$$



2021.02.02

$$\varphi^{(l)}(\underline{x}) = \begin{bmatrix} \varphi^{(l-1)}(\underline{x}) \\ \hline \varphi_{\cdot}^{l}(\underline{x}) \end{bmatrix} \qquad \underline{w}^{*} = \left(\underline{\underline{\Phi}}^{T}\underline{\underline{\Phi}}\right)^{-1}\underline{\underline{\Phi}}^{T}\underline{d} \qquad \underline{\underline{\Phi}}^{(l)} = \begin{bmatrix} | & | & & | \\ \varphi_1 & \varphi_2 & \dots & \varphi_l \\ | & | & & | \end{bmatrix}$$

$$\underline{w}^{*(l)} = \arg\min_{\underline{w}^{(l)}} \left\{ \min_{i_1, i_2, \dots, i_l} \left\| \begin{bmatrix} | & | & & | \\ \varphi_{i_1} & \varphi_{i_2} & \dots & \varphi_{i_l} \\ | & | & & | \end{bmatrix} \cdot \underline{w} - \underline{d} \right\|_2^2 \right\}$$

$$\underline{\varphi}_{i_1} = \begin{bmatrix} \exp(-(1/2\sigma)^2 \cdot \|\underline{x}_1 - \underline{c}_{i_1}\|_2^2) \\ \vdots \\ \exp(-(1/2\sigma)^2 \|\underline{x}_n - \underline{c}_{i_1}\|_2^2) \end{bmatrix}$$

Orthogonal Least Squares:

$$\underline{\underline{\Phi}}' = \begin{bmatrix} | & | & & | \\ \varphi_1' & \varphi_2' & \dots & \varphi_l' \\ | & | & & | \end{bmatrix} \qquad \underline{\underline{\Phi}}'^{T} \cdot \underline{\underline{\Phi}}' = \cancel{\underline{\underline{I}}} \begin{bmatrix} \ddots & 0 \\ 0 & \ddots \end{bmatrix}$$

$k = 0$: $\quad \underline{d}^{(0)} = \underline{d} \qquad \Gamma_0 = \{1, 2, \dots, k\} \qquad err(0) = \|\underline{d}\|_2^2 \qquad \underline{w}^{(0)} = [\,]$

$k > 0$

$k > 0$. iterációnál:

$$w^{*}_{j^{(k)}} = \arg\min_{w'} \Big\{ \underbrace{\min_{j \in I^{(k-1)}} \{ \| \underline{d}^{(k-1)} - \underline{g}^{j^{(k)}} \cdot w' \|_2^2 \}}_{j^{(k)} = \arg\min \{ - \| - \}} \Big\}$$

$$\arg\min_{w'} \{ \underbrace{\| \underline{d}^{(k-1)} - \underline{g}^{j^{(k)}} \cdot w' \|_2^2}_{w'^2 \cdot \underline{g}^{j^{(k)}T} \underline{g}^{j^{(k)}} - 2 w' \cdot \underline{d}^{(k-1)T} \cdot \underline{g}^{j^{(k)}}} \} = \frac{\underline{g}^{j^{(k)}T} \cdot \underline{d}^{(k-1)}}{\underline{g}^{j^{(k)}T} \cdot \underline{g}^{j^{(k)}}} = w^{*}_{j^{(k)}}$$

- $\underline{w}^{(k)} = \begin{bmatrix} \underline{w}^{(k-1)} \\ w^{*}_{j^{(k)}} \end{bmatrix}$; $\underline{d}^{(k)} := \underline{d}^{(k-1)} - \underline{g}^{j^{(k)}} \cdot w^{*}_{j^{(k)}}$; $\underline{I}^{(k)} := \underline{I}^{(k-1)} \setminus \{j^{(k)}\}$
- $\forall i \in I^{(k)}$: $\quad b_{i,j^{(k)}} = \dfrac{\underline{g}_i^{(k)T} \cdot \underline{g}_{j^{(k)}}}{\underline{g}_{j^{(k)}}^T \cdot \underline{g}_{j^{(k)}}}$; $\underline{g}_i^{(k)} := \underline{g}_i - b_{i,j^{(k)}} \cdot \underline{g}^{j^{(k)}}$

$$err^{(k)} = \| \underline{d}^{(k)} \|_2^2$$

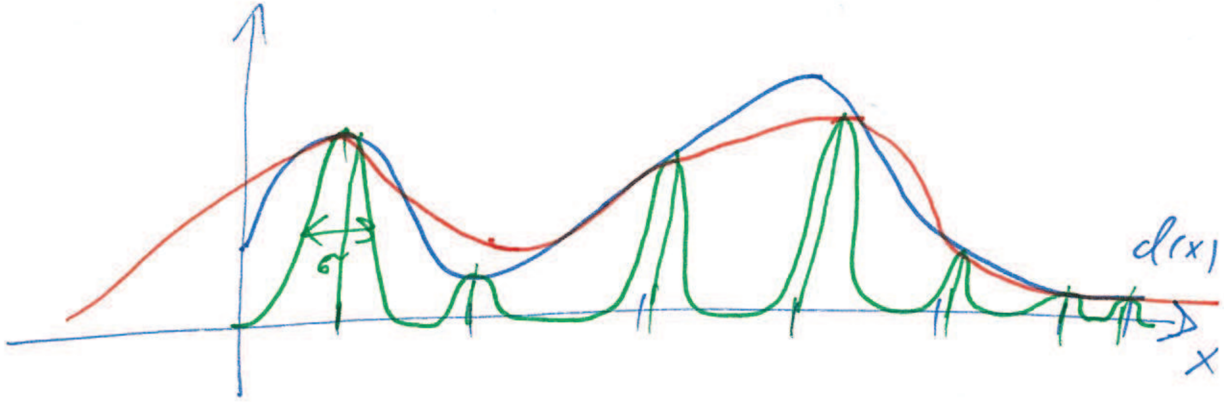
$$\underline{\Phi} \cdot \underbrace{\begin{array}{c} \\ \\ \\ \varphi\, j(1) \end{array}\begin{bmatrix} 0 & 1 & \cdots & 0 & & & \\ \vdots & 0 & & & & & \\ \vdots & \vdots & & & & & \\ 0 & 0 & & 1 & & & \\ 1 & -b_{1,j(1)} & \cdots & -b_{j(1),j(1)} & -b_{j(1)+1,j(1)} & \cdots & -b_{n,j(1)} \\ 0 & 0 & & 0 & 1 & & \\ \vdots & \vdots & & \vdots & 0 & \ddots & \\ 0 & 0 & & 0 & 0 & & 1 \end{bmatrix} \cdot \begin{array}{c} j(2) \\ \begin{bmatrix} 1 & 0 \\ 0 & j(2) \\ \vdots & 0 \\ \vdots & \vdots \\ 0 & 0 \end{bmatrix} \end{array} \begin{bmatrix} \cdots \end{bmatrix} \cdot \ldots}_{\underline{W}'^{(k)}} \cdot \underline{W}^{(q)}$$

$$g(x) = \sum_i \underline{\varphi}^{(i)}(\underline{x}) \cdot w'^{(k)}_i$$

<u>K-means:</u>

$$\underset{\underline{c}_1, \underline{c}_2, \ldots, \underline{c}_k}{\arg\min} \left\{ \sum_{i=1}^{n} \cancel{x_i} \min_{j \in \{1,\ldots,k\}} \left\{ \|\underline{x}_i - \underline{c}_j\|_2^2 \right\} \right\}$$

σ ~~átl~~ megválasztása:



$$\sigma = \frac{1}{n} \cdot \sum_i \min_k \{ \|\underline{x}_i - \underline{c}_k\|_2^2 \}$$

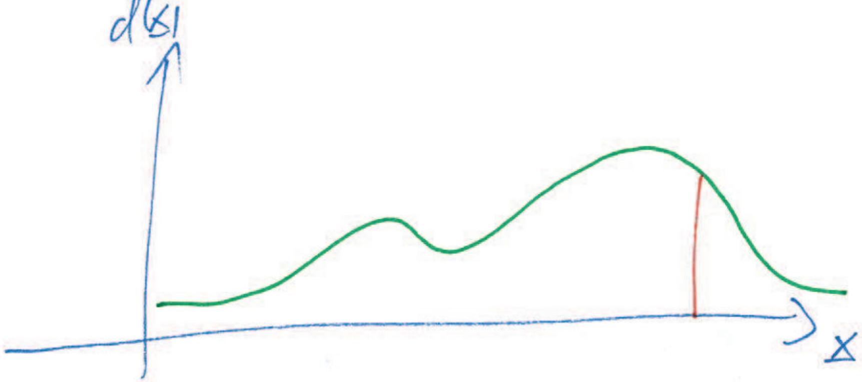
Bázisfgv-ek lokális detektálása:

– $\forall i \; \int_x (\varphi_i(\underline{x})^2) < \infty \qquad \exists c_i : \; \varphi_i(c_i) = 1 \qquad \forall \|\underline{e}\|_2 = 1$

$$\lim_{t \to \infty} \varphi_i(c_i + \underline{e} \cdot t) = 0$$

ezáltal univerzális approx.

– lokálisan általánosít:



– Online tanulásnál nem felejt

Kernel gépek

$\dim\{\underline{\varphi}(x)\} > N$

$$\underline{\underline{\Phi}} = \begin{bmatrix} - \underline{\varphi}(x_1)^T - \\ \vdots \\ - \underline{\varphi}(x_N)^T - \end{bmatrix}$$

$$\underline{w}^* = \arg\min_{\underline{w}} \{ \| \underline{d} - \underline{\underline{\Phi}} \cdot \underline{w} \|_2 \}$$

$$\underline{w}^* = \underbrace{\underline{\underline{\Phi}}^T \overbrace{\left(\underline{\underline{\Phi}} \cdot \underline{\underline{\Phi}}^T \right)^{-1}}^{\underline{\alpha}}}_{\underline{\underline{\Phi}}^+} \underline{d}$$

$$y(x) = \underline{\varphi}(x)^T \cdot \underline{w}^*$$

$$y(x) = \sum_i \underbrace{\underline{\varphi}(x)^T \cdot \underline{\varphi}(\underline{x}_i)}_{K(\underline{x}, \underline{x}_i) = \underline{\varphi}(x)^T \cdot \underline{\varphi}(\underline{x}_i)} \cdot \alpha_i$$