2021. 03. 16 – 18.

Statisztikus tanuláselmélet

Kockázat:
$$R(\underline{w}) = \underset{(\underline{x},d)\sim\mathcal{X}}{\mathbb{E}} \left\{ L(d, f(\underline{x}, \underline{w})) \right\}$$

$$R_{emp}(\underline{w}) = \frac{1}{|S|} \cdot \sum_{i=1}^{|S|} L(d_i, f(\underline{x}_i, \underline{w}))$$

pl.: $- L(d, y) = (y-d)^2 \quad ; \quad \max\{|y-d|-\varepsilon, 0\} \quad ; \quad |y-d|$

$- L(d, y) = -d \cdot \log(y) - (1-d)\log(1-y)$

$$= \sum_c -d_c \cdot \log(y_c) \qquad d_i = P\{\underline{x} \in i. \text{ osztály}\}$$
$$y_i = P_f\{\underline{x} \in i. \text{ osztály}\}$$

Torzítás variancia dilemma:

$$R(\underline{w}) = \underset{(\underline{x},d)\sim\mathcal{X}}{\mathbb{E}} \left\{ (d - f(\underline{x}, \underline{w}))^2 \right\}$$

$d = f(\underline{x}) + \varepsilon$

1. $\mathbb{E}\{\varepsilon\} = 0$ $\quad \varepsilon$: i.i.d

$\hat{f}_S(\underline{x})$: S halmaz alapján $f(\cdot)$ becslője

$$\mathbb{E}_{\underline{x}}\{f(\underline{x})\} = \int_{\underline{x}} p(\underline{x})\cdot f(\underline{x})\,d\underline{x}$$

$$R = \mathbb{E}_{\varepsilon, S, \underline{x} \mid \mathcal{X}}\left\{\left(f(\underline{x}) + \varepsilon - \hat{f}_S(\underline{x})\right)^2\right\} = \mathbb{E}\Big\{ \varepsilon^2 + 2\cdot\varepsilon\left(f(\underline{x}) - \hat{f}_S(\underline{x})\right) +$$

$$+ \left(f(\underline{x}) - \hat{f}_S(\underline{x})\right)^2\Big\} = \underbrace{\mathbb{E}_{\varepsilon}\{\varepsilon^2\}}_{\mathrm{Var}\{\varepsilon\}} + 2\cdot \underbrace{\mathbb{E}_{\varepsilon}\{\varepsilon\}}_{0}\cdot \mathbb{E}_{\underline{x}, S}\left\{\left(f(\underline{x}) - \hat{f}_S(\underline{x})\right)\right\}$$

$$+ \mathbb{E}\left\{\left(f(\underline{x}) - \hat{f}_S(\underline{x})\right)^2\right\} = \mathrm{Var}\{\varepsilon\} + \mathbb{E}_{\underline{x}, S}\left\{\left(f(\underline{x}) - \hat{f}_S(\underline{x})\right)^2\right\}$$

$$\mathbb{E}_{\underline{x}, S \mid \mathcal{X}}\left\{\left(f(\underline{x}) - \mathbb{E}_S\{\hat{f}_S(\underline{x})\} + \mathbb{E}_S\{\hat{f}_S(\underline{x})\} - \hat{f}_S(\underline{x})\right)^2\right\} =$$

$$= \mathbb{E}\left\{\left(f(\underline{x}) - \mathbb{E}_S\{\hat{f}_S(\underline{x})\}\right)^2\right\} + \mathbb{E}\left\{\left(\mathbb{E}_S\{\hat{f}_S(\underline{x})\} - \hat{f}_S(\underline{x})\right)^2\right\} + (2)$$

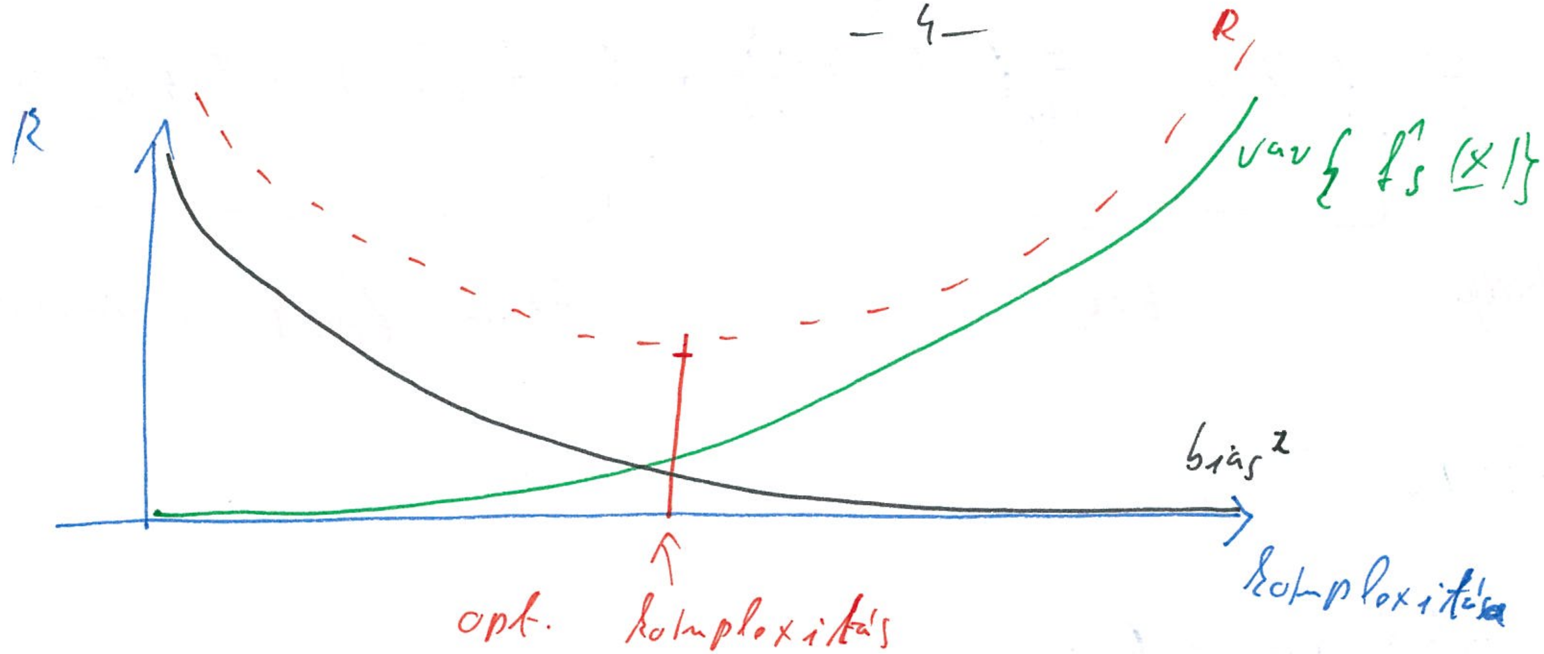
$$+ 2\cdot \mathbb{E}\left\{\left(f(\underline{x}) - \mathbb{E}_S\{\hat{f}_S(\underline{x})\}\right)\cdot \underbrace{\left(\mathbb{E}_S\{\hat{f}_S(\underline{x})\} - \hat{f}_S(\underline{x})\right)}_{0}\right\}$$

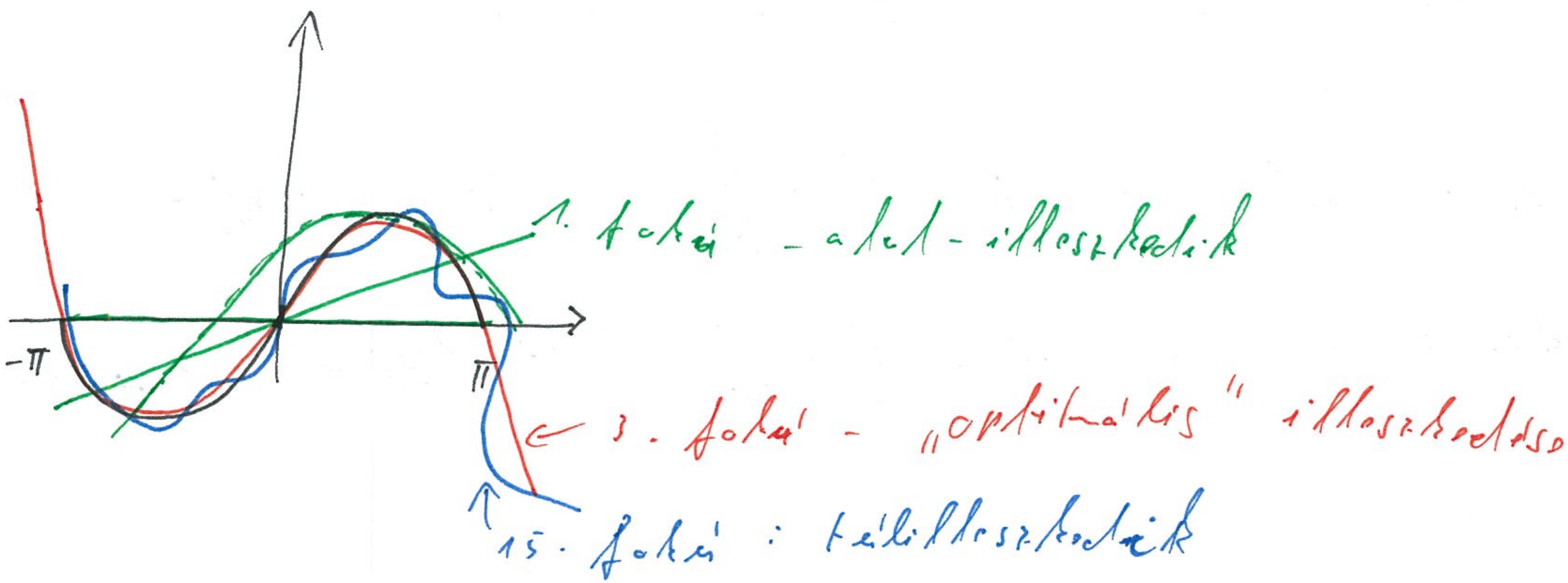
$$\int_{\underline{x}}\int_{S} g(\underline{x})\cdot h(S, \underline{x})\cdot p(S)\cdot p(\underline{x})\,dS\,d\underline{x}$$

$$R = \underbrace{\mathrm{var}\{\varepsilon\}}_{\text{megf. zaj variációja}} + \underbrace{\mathop{E}_{\underline{x},S|\chi}\left\{\left(f(\underline{x}) - \mathop{E}_{S|\chi}\left\{\hat{f}_S(\underline{x})\right\}\right)^2\right\}}_{\text{bias}^2} + \underbrace{\mathop{E}_{\underline{x},S|\chi}\left\{\left(\hat{f}_S(\underline{x}) - E\left\{\hat{f}_S(\underline{x})\right\}\right)^2\right\}}_{\text{becslő } (\hat{f}_S(\underline{x})) \text{ variációja}}$$

- Hipotézistér ($\bigcup_{\underline{w}} f(\underline{x},\underline{w})$) mérete / $\hat{f}_S(\underline{x})$ $f(\underline{x},\underline{w})$ szabadságfoka
- $|S|$ – adat augmentáció
- tanító eljárás feladat $\hat{f}_S(\underline{x})$ kiválasztása

- torzítatlanság: $\underline{w}^* = \arg\min_{\underline{w}} \left\{ \frac{1}{|S|} \cdot \sum_{i=1}^{|S|} L(d_i, f(\underline{x}_i, \underline{w})) \right\}$
- regularizáció: $\underline{w}^* = \arg\min_{\underline{w}} \left\{ \frac{1}{|S|} \cdot \sum_i L(d_i, f(\underline{x}_i, \underline{w})) + \lambda \cdot \Omega(\underline{w}) \right\}$


R
R
var{ $\hat{f}_S(\underline{x})$ }
bias²
komplexitás
opt. komplexitás


−π
π
1. fokú – alul-illeszkedik
3. fokú – „optimális" illeszkedés
15. fokú : túlilleszkedik

regularizáció – $\hat{f}_S(\underline{x})$ variációjának a csökkentése

- pl. (SVM-c)

min. $\underline{w}^T \underline{w}/2 + C \cdot \sum_i \xi_i$ ← Tikhonov [red] Tikhonov

s. t. $(\underline{w}^T \underline{x}_i + b)\, d_i \geq 1 - \xi_i$

$\xi_i \geq 0$

- ha $C = \infty$: – tanítópontokat illesztjük jól
  – $\|\underline{w}\|_2^2$ minimális

- ha $C < \infty$: – bias² nől
  – $\mathrm{var}\{\hat{f}_S(\underline{x})\} \downarrow$

- regressziós eset – MLP:

$$y(\underline{x} + \underline{\Delta x}) = y(\underline{x}) + \underline{\Delta x}^T \cdot \underline{\nabla y(\underline{x})} + \cancel{O(\|\underline{\Delta x}\|_2^2)}\ O(\Delta x^2)$$

$$\max_{\underline{x}, \underline{x}'} \{ |y(\underline{x}) - y(\underline{x}')| / \|\underline{x} - \underline{x}'\|_2 \}$$

– bias²: ↑
– $\mathrm{var}\{\hat{f}_S\}$: ↓

↑ csökkenti : $C = R_{\mathrm{emp}}(\underline{w}) + \lambda \cdot \|\underline{w}\|_2^2$

Vapnik - Chervonenkis statisztikus tanuláselmélet

Вапник - Червоненкис (VC / ВЧ)

ERM (Empirical Risk Minimization) konzisztenciája:

$$\lim_{l \to \infty} R_{emp}(\underline{w}^*/l) \to \lim_{l \to \infty} R(\underline{w}^*/l) = R(\underline{w}^*) \qquad \underline{w}^* = \arg\min_{\underline{w}} \{R(\underline{w})\}$$

- Tétel: ERM konzisztens a.cs.a.

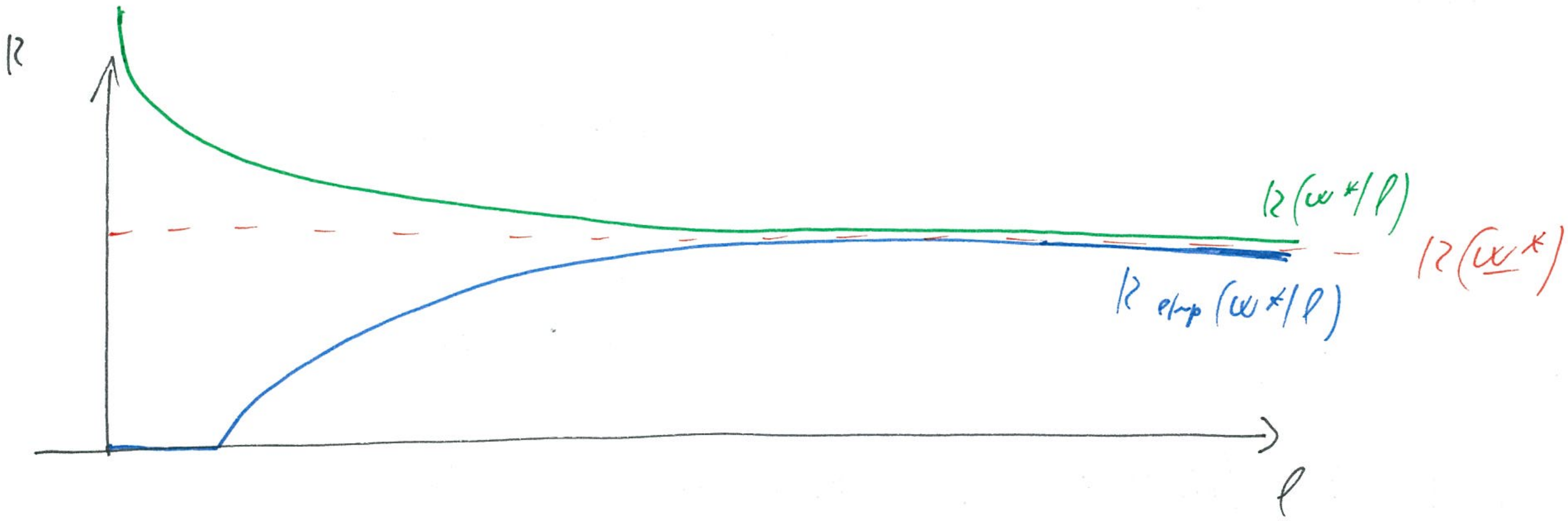
$$\lim_{l \to \infty} \{P\{\sup_{\underline{w}} \{ |R_{emp}(\underline{w}) - R(\underline{w})| \} > \varepsilon\}\} = 0 \qquad \varepsilon > 0$$

⊖ - Nem konstruktív
 - konvergencia sebessége nem ad érdemleges infót.

- Tétel ERM gyorsan konvergál:

$$P\{ R(\underline{w}^*/l) - R(\underline{w}^*) > \varepsilon \} < \exp(-c \cdot l \cdot \varepsilon^2)$$

$$\forall \varepsilon > 0 \quad \forall l \geq l_0$$



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

~~ind~~ I

- indikátor fgv $\quad I(f(\cdot, \underline{w}); \underline{z}) = \begin{cases} 1 / & f(\underline{z}, \underline{w}) > 0 \\ 0 / & \text{egyébként} \end{cases}$

- $Z_\ell = \{\underline{z}_1, \underline{z}_2, \ldots, \underline{z}_\ell\}$

– f diverzitása: $N_f(Z_\ell) := \#\Big| \bigcup_{\underline{w}} \underbrace{\times_{k} I(f(\cdot|\underline{w}), \underline{z}_k)}_{\begin{bmatrix} I(f(\cdot|\underline{w}), \underline{z}_1) \\ I(f(\cdot|\underline{w}), \underline{z}_2) \\ \vdots \\ I(f(\cdot|\underline{w}), \underline{z}_\ell) \end{bmatrix}} \Big|$

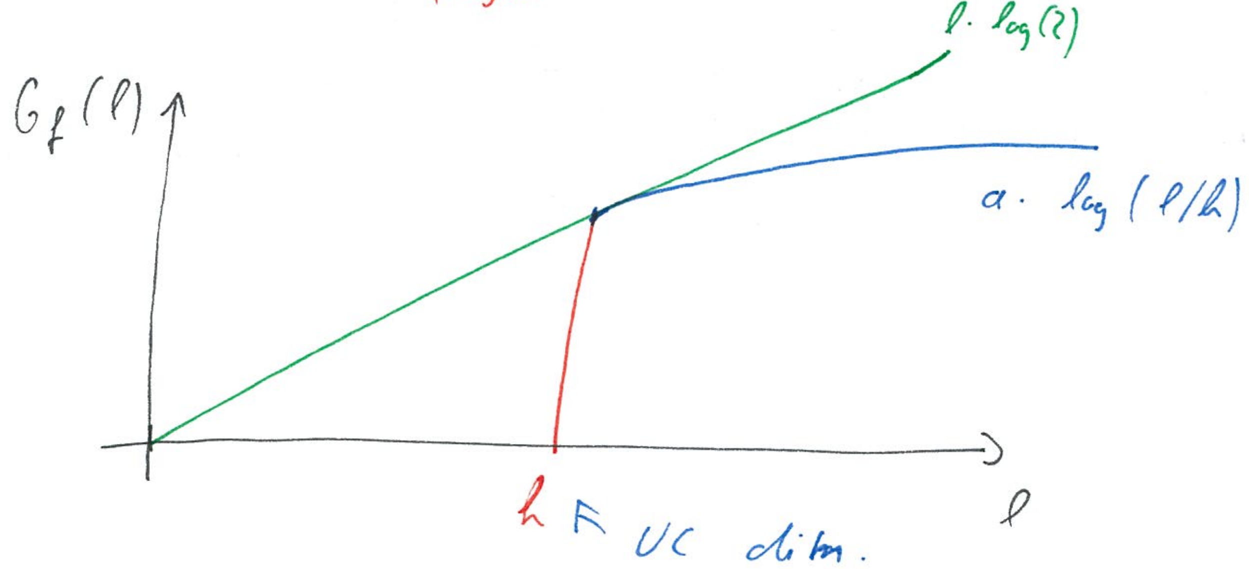
– f entrópiája Z_ℓ felett: $H_f(Z_\ell) := \log(N_f(Z_\ell))$

– f VC entrópiája: $\cancel{H_f(Z_\ell)}$ $H_f(\ell) := \mathbb{E}_{\substack{Z_\ell \sim \mathcal{X} \\ |Z_\ell| = \ell}} \{H_f(Z_\ell)\}$

ERM konzisztens $\Leftrightarrow$ $\lim_{\ell \to \infty} H_f(\ell)/\ell = 0$

– Növekedési fgv: $G_f(\ell) := \log\{\max_{|Z_\ell| = \ell}\{N_f(Z_\ell)\}\}$

ERM konzisztens $\Leftarrow \lim_{l \to \infty} G_f(l)/l = 0$



f VC dim.-ja $h \iff \exists \{\underline{z}_1, \underline{z}_2, \dots, \underline{z}_h\} = Z_h : N_f(Z_h) = 2^h$

$\nexists \{\underline{z}_1, \underline{z}_2, \dots, \underline{z}_{h+1}\} = Z_{h+1} : N_f(Z_{h+1}) = 2^{h+1}$

pl.: f lin. fgv-ek osztálya $f(\underline{x}, \underline{w}) = \underline{w}^T \underline{x} + b \quad \underline{x} \in \mathbb{R}^N$

$h = N+1$

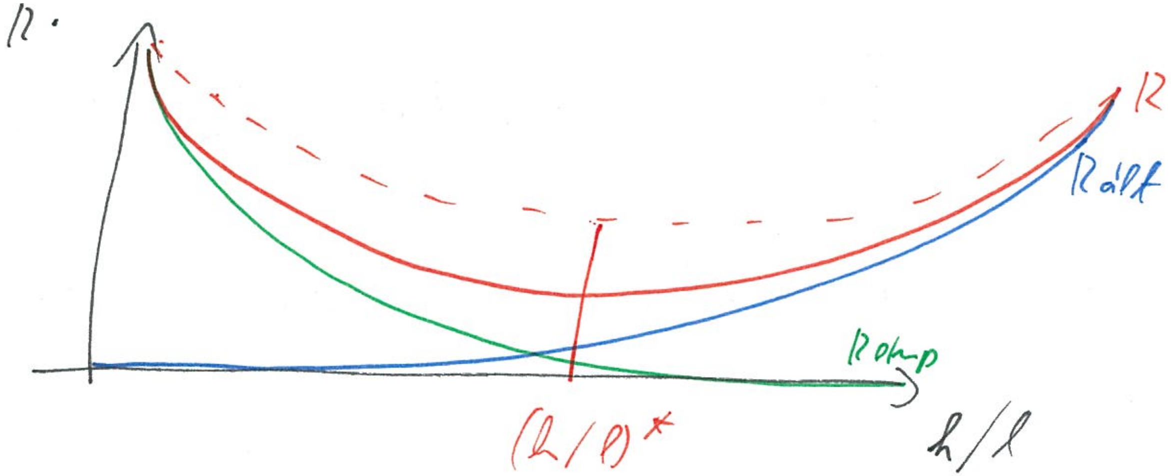
$$P\left\{ R(\underline{w}) \le R_{emp}(\underline{w}) + \frac{\varepsilon(h)}{2}\sqrt{1+\frac{4\,R_{emp}(\underline{w})}{\varepsilon(h)}} \right\} \ge 1-\eta$$

äll. Luibentug

$$\varepsilon(h) := 4 \cdot \frac{h\left(\ln(2\,l/h)+1\right) - \ln(\eta/4)}{l}$$

$$\varepsilon(h) \sim h/l$$

äll. Luiba $\quad R(\underline{w}) - R_{emp}(\underline{w}) \quad \sim \quad (h/l)$

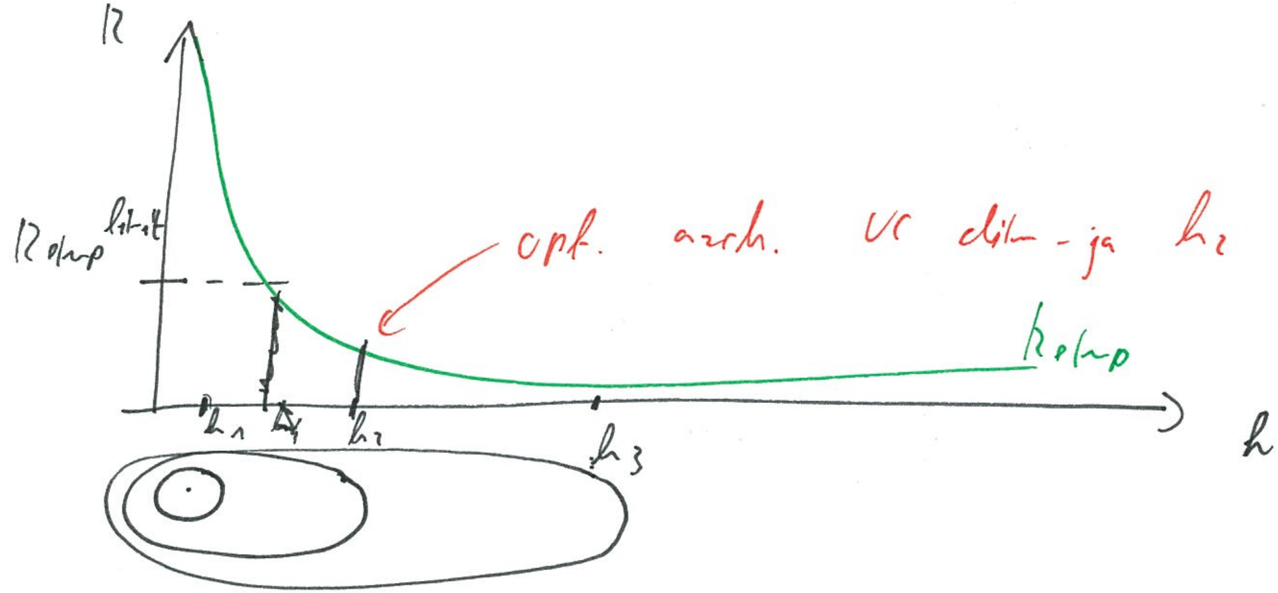


SVM-c

$$h \le \min\{R^2\|\underline{w}\|_2^2 ; N+1\}$$ ⇐ SVM minimalizálja strukturálisan R

$$R^{H} = (1/2)\sqrt{\max_{\underline{\varphi}_1,\underline{\varphi}_2}\{\|\underline{\varphi}_1 - \underline{\varphi}_2\|_2^2\}}$$

Strukturális Kockázat Minimalizálás (SRM):



$$\bigcup_{w} f(\cdot, \underline{w})$$

Tanulás mint statisztikai becslés:

$$\begin{tikzcd} & & \varepsilon \arrow[d] & \\ \underline{x} \arrow[r] & \boxed{f(\cdot,\underline{w})} \arrow[r, "y^*"] & \oplus \arrow[r] & d \end{tikzcd}$$

$\underline{x}_1, \; d_1$

$\underline{x}_2, \; d_2$

$\vdots$

1./ Maximum likelihood becslés:

$$\varepsilon \sim \eta \qquad p(\varepsilon_i, \varepsilon_j) = p(\varepsilon_i) \cdot p(\varepsilon_j)$$

$$\mathcal{L} = \prod_{i=1}^{L} \underbrace{p_\eta(d_i \,|\, f(\cdot, \underline{w}), \underline{x}_i)}_{p_\eta(\underline{d}_i \,|\, \underline{w}, \underline{x}_i)}$$

$$\underline{w}^* = \arg\max_{\underline{w}} \{\mathcal{L}\} = \arg\min_{\underline{w}} \{ \overbrace{-\log(\mathcal{L})}^{\text{hűség tag}} \}$$

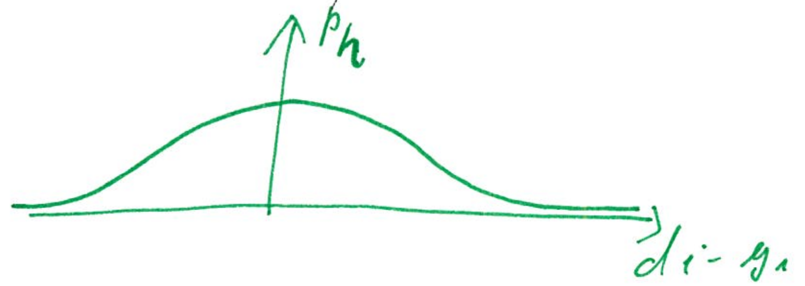
ERM:

$$C = -\log(\mathcal{L}) = -\sum_i \log(p_\eta(d_i \,|\, \underline{w}, \underline{x}_i)) = \sum_i L(d_i, f(\underline{x}_i, \underline{w}))$$

Pl.: $- \; L(d,y) = (y-d)^2$: $\quad \varepsilon \sim \mathcal{N}(\underline{0}, \sigma^2 \cdot \underline{\underline{I}})$

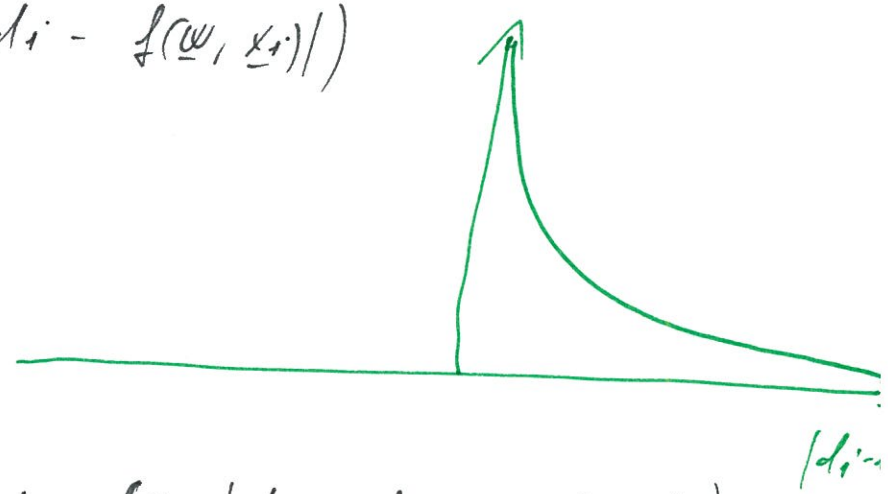
$$p_h(d_i \mid \underline{w}, \underline{x_i}) = \frac{1}{\sqrt{2\pi}\cdot\sigma} \cdot \exp\left(-\frac{1}{2\sigma^2}\cdot\left(d_i - f(\underline{w}, \underline{x_i})\right)^2\right)$$

$$-\log(p_h) = K + a\cdot\left(d_i - \underbrace{f(\underline{w}, \underline{x_i})}_{y_i}\right)^2$$



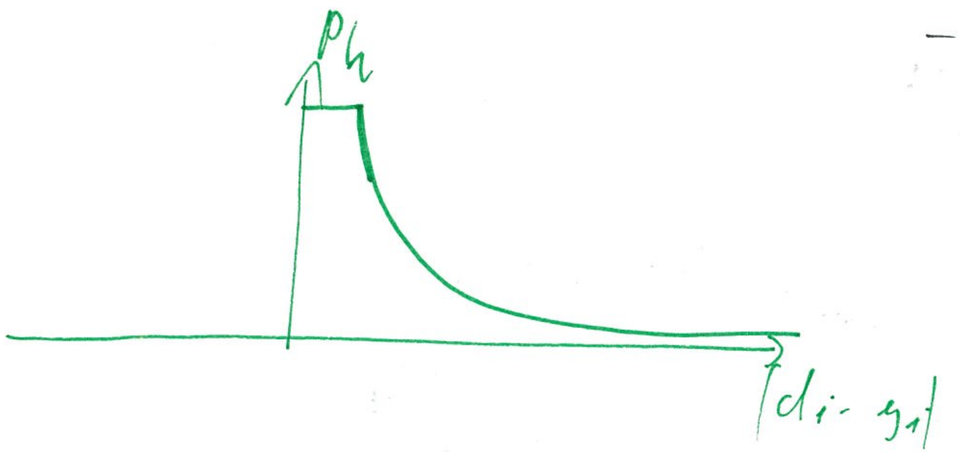
$- \; L(d,y) = |d-y| \qquad \varepsilon \sim \text{Laplace}(\underline{0}, b)$

$$p_h(d_i \mid \underline{w}, \underline{x_i}) = C \cdot \exp\left(-(1/b)\cdot|d_i - f(\underline{w}, \underline{x_i})|\right)$$



$- \; L(d,y) = \max\{0, |d-y| - \varepsilon\}$

$$p_h(d_i \mid \underline{w}, \underline{x_i}) = C \cdot \exp\left(-(1/b)\cdot \max\{0, |d_i - f(\underline{w}, \underline{x_i})| - \varepsilon\}\right)$$


p_k
$|d_i - y_1|$