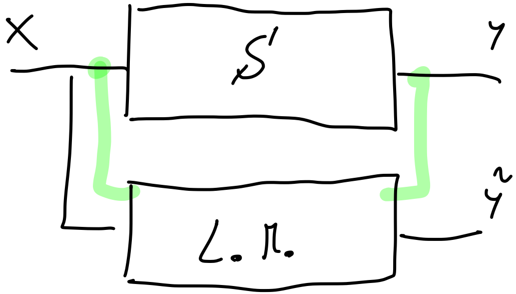


$\tilde{y} = y$

$D_n = \{(x_1, y_1), \dots, (x_n, y_n)\}$ i. i. d. $x \in \mathbb{R}$ $y \in \mathbb{R}$

$$f(x) = \sum_{i=0}^{p} c_i x^i$$

$$\ell(f(x), y) = (y - f(x))^2$$

$$C(f) = \begin{cases} \|c\|^2 \\ \int \left| \frac{d^2 f}{dx^2} \right|^2 dx \end{cases}$$

$$\boxed{\min_f \underbrace{\overset{\text{STAT.}}{\hat{R}(f)} + \overset{\text{LEARNING}}{\lambda C(f)}}_{R(f)}}$$

$\underline{x} \in \mathbb{R}^d$

- $\overset{1}{c_0}$
- $// + \underbrace{c_1 x_1 + \dots + c_d x_d}_{}$ $\quad d$
- $// + // + \underbrace{\bar{c}_1 x_1^2 + \dots + \bar{c}_d x_d^2}_{d} + \underbrace{\dots x_1 x_2 + \dots + \dots x_{d-1} x_d}_{\binom{d}{2} \simeq d^2}$

$p \qquad p \qquad p$

- $\binom{d}{p} \simeq d^p$

$d = 10^6 \qquad p = 6 \quad \rightarrow \quad (10^6)^6 \simeq (2^{16})^6 \simeq 2^{100}$

$$f(\underline{x}) = \underline{w} \cdot \underline{x}$$

$$X = \begin{bmatrix} \underline{x}_1' \\ \vdots \\ \underline{x}_m' \end{bmatrix} \in \mathbb{R}^{m \times d} \qquad \underline{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$$

$$\ell(f(\underline{x}), y) = (y - f(x))^2$$

$$C(f) = \|\underline{w}\|^2$$

$$\min_{\underline{w}} \sum_{i=1}^{m} (\underline{w}\, \underline{x}_i - y_i)^2 + \lambda \|\underline{w}\|^2$$

$$\min_{\underline{w}} \|X\underline{w} - \underline{y}\|^2 + \lambda \|\underline{w}\|^2$$

RIDGE REGRESSION

$$\nabla_w \left(\underline{w}' X' X \underline{w} - 2\, \underline{w}' X' \underline{y} + \underline{y}' \underline{y} + \lambda\, \underline{w}' I \underline{w} \right) = \underline{0}$$

$$\cancel{2} X' X \underline{w} - \cancel{2} X' \underline{y} + \cancel{2} \lambda I \underline{w} = \underline{0}$$

$$(X' X + \lambda I) \underline{w} = X' \underline{y} \qquad [\;]_{d \times d} [\;]_{d} = [\;]_{d}$$

$$\min_{\underline{w}} \|X\underline{w} - \underline{y}\|^2 + \lambda \|\underline{w}\|^2$$

$$A\underline{x} = b$$

$$\underline{w}^* = \underline{w}_\perp + \underline{w}_\parallel$$

X

$$\underline{w}_\parallel = X'\underline{\alpha} = \sum_{i=1}^{n} \alpha_i \underline{x}_i$$

$$X = \begin{bmatrix} \underline{x}_1' \\ \vdots \\ \underline{x}_n' \end{bmatrix}$$

$$\|X(\underline{w}_\parallel + \underline{w}_\perp) - y\|^2 + \lambda\|\underline{w}_\parallel + \underline{w}_\perp\|^2$$

$$\underbrace{\|X\underline{w}_\parallel + \underbrace{X\underline{w}_\perp}_{=0} - y\|^2}_{} + \underbrace{\lambda\|\underline{w}_\parallel\|^2}_{\geq 0} + \underbrace{\cancel{\lambda\|\underline{w}_\perp\|^2}}_{\geq 0}$$

REP.
TH.

$\underline{w}_\parallel$?

$$\underline{w}^* = \underline{w}_\parallel$$

$$\min_{\underline{w}} \; \| X\underline{w} - \underline{y} \|^2 + \lambda \|\underline{w}\|^2$$

$$\underbrace{[\quad]}_{d\times d}\;\underbrace{[\;]}_{d} = \underbrace{[\;]}_{d}$$

$$\underline{w}^* = X' \underline{\alpha}$$

$$\min_{\underline{\alpha}} \; \| \underbrace{XX'}_{Q} \underline{\alpha} - y \|^2 + \lambda \, \underline{\alpha}' \underbrace{XX'}_{} \underline{\alpha}$$

$$Q \to Q_{ij} = \underline{x}_i' \underline{x}_j = Q_{ji}$$

$$\nabla_{\underline{\alpha}} \left(\underline{\alpha}' Q' Q \underline{\alpha} - 2 \alpha' Q' \underline{y} + \underline{y}' \underline{y} + \lambda \underline{\alpha}' Q \underline{\alpha} \right) = \underline{0}$$

$$\cancel{2}\cancel{Q'}\, Q \underline{\alpha} - \cancel{2}\cancel{Q'}\, \underline{y} + \cancel{2} \lambda \cancel{Q'}\, \underline{\alpha} = \underline{0}$$

$Q \succ 0$

$$(Q + \lambda I)\, \underline{\alpha} = \underline{y}$$

$$\begin{array}{cc} X & X' \\ n\times d & d\times n \end{array}$$

$n \times n$

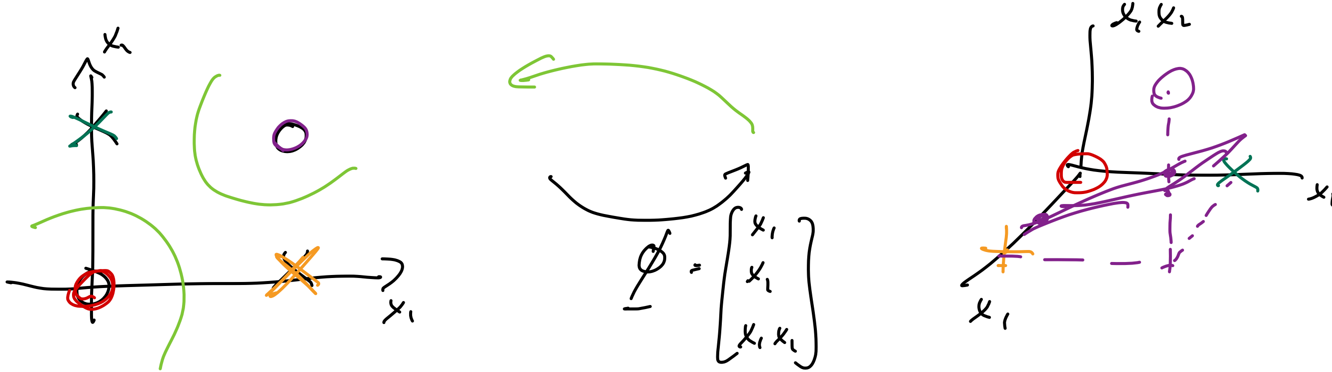
$$\underbrace{[\;]}_{n}$$

$$n\underbrace{[\quad]}_{n}\;\underbrace{[\;]}_{n} = \underbrace{[\;]}_{n}$$

$$f(\underline{x}) = \underline{w}'\underline{x} = \underline{\alpha}' X \underline{x} = \sum_{i=1}^{m} \alpha_i \underbrace{\underline{x}_i'\,\underline{x}}_{\underline{\phi}'(\underline{x}_i)\,\phi(\underline{x}) \;\to\; K(x_i,\, x)}$$

$$\underline{\alpha} : \quad (Q + \lambda I)\,\underline{\alpha} = \underline{y}$$

$$\hookrightarrow Q_{ik} = \underbrace{\underline{x}_i'\,\underline{x}_k}_{\underline{\phi}'(x_i)\,\underline{\phi}'(\underline{x}_k) \;\to\; K(\underline{x}_i,\, \underline{x}_k)}$$



$$\underline{w}\;\underline{\phi}(\underline{x})$$

MERCER $\rightarrow$ KERNELS $\quad \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}$

GAUSSIAN RBF $\rightarrow$ $e^{-\gamma \|\underline{x}_i - \underline{x}_j\|}$, $\langle \underline{x}_i , \underline{x}_j \rangle^P$

$\uparrow$ $\quad \mathbb{R}$

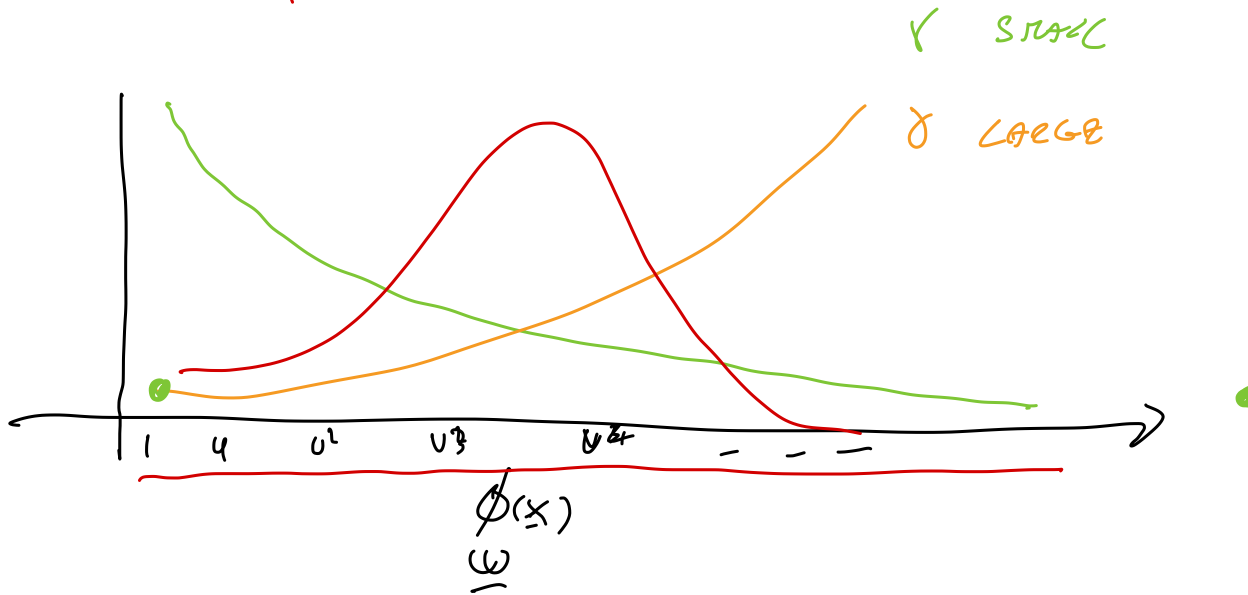
$\underline{\phi}'(x_i) \underline{\phi}(\underline{x}_j)$

$e^{-\gamma (u-v)^2}$ $\qquad d=1$ $\qquad e^x = \sum_{i=0}^{\infty} \frac{1}{i!} x^i$

$= e^{-\gamma u^2} e^{-\gamma v^2} e^{2\gamma uv}$

$= e^{-\gamma u^2} e^{-\gamma v^2} \sum_{i=0}^{\infty} \frac{1}{i!} (2\gamma uv)^i = e^{-\gamma u^2} \begin{bmatrix} 1 \\ \sqrt{2\gamma}\, u \\ 2\gamma\, u^2 \\ \vdots \end{bmatrix}' e^{-\gamma v^2} \begin{bmatrix} 1 \\ \sqrt{2\gamma}\, v \\ 2\gamma\, v^2 \\ \vdots \end{bmatrix} = \underline{\phi}'(u)\, \underline{\phi}(\underline{v})$

$$\min_{\underline{w}} \; \|X\underline{w} - \underline{y}\|^2 + \lambda \|\underline{w}\|^2$$



$$D_n = \{ (\underline{x}_1, y_1), \dots, (\underline{x}_n, y_n) \}$$

$$f(\underline{x}) = \underline{w}\, \underline{\phi}(\underline{x}) \qquad K(\cdot, \cdot), \gamma$$

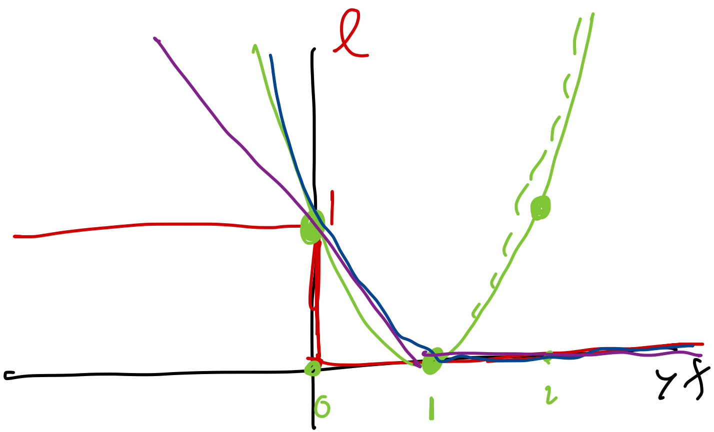
$$\ell(f(\underline{x}), y) = (y - f(\underline{x}))^2$$

$$C(f) = \|\underline{w}\|^2$$

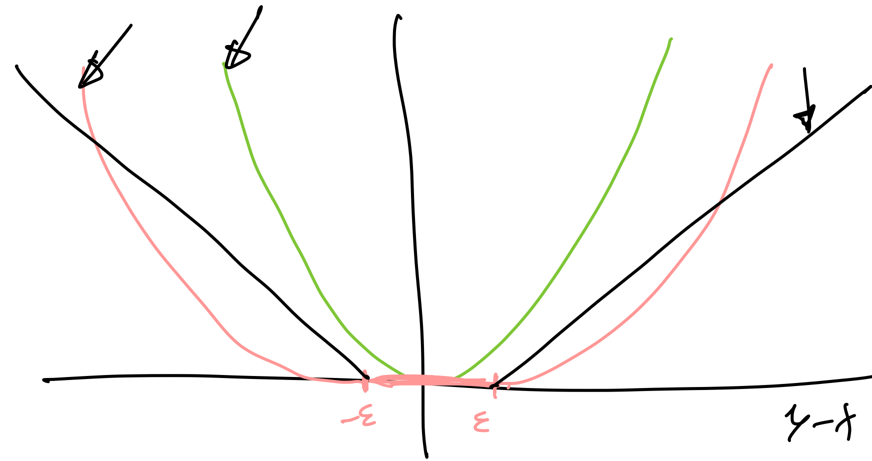
$$\min_{f} \hat{R}(f) + \lambda C(f)$$

CLASS.

$y \in \{-1, 1\}$

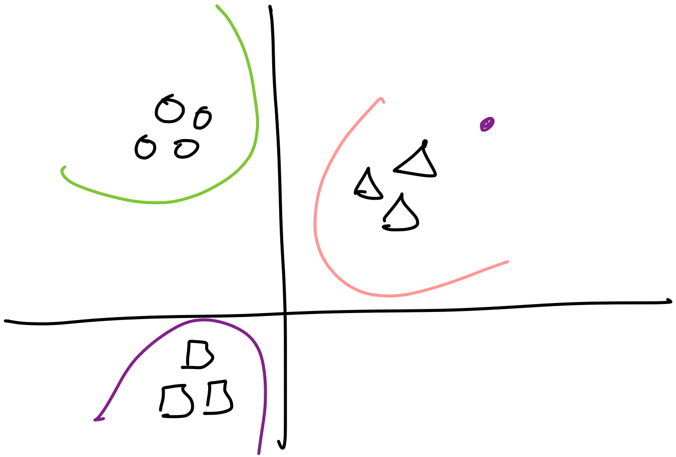


REGRESS



CLASS

$y = \{1, 2, 3, \ldots, c\}$



OVA

- C B.C. P.
- m SAMPLES
- CONFIDENCE
- UNBALANCED

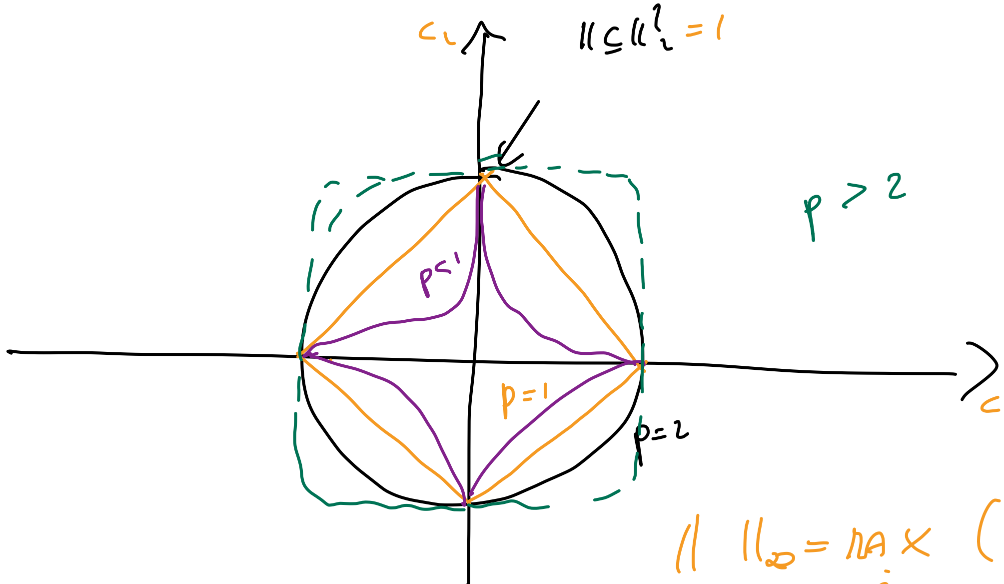
AVA

- $\binom{c}{2} \simeq c^2$
- $2 \frac{m}{c}$
- M. V.
- BALANCED

$$L(h) = \|c\|_p = \sqrt[p]{\sum_{i=1}^{d} |c_i|^{p}}$$

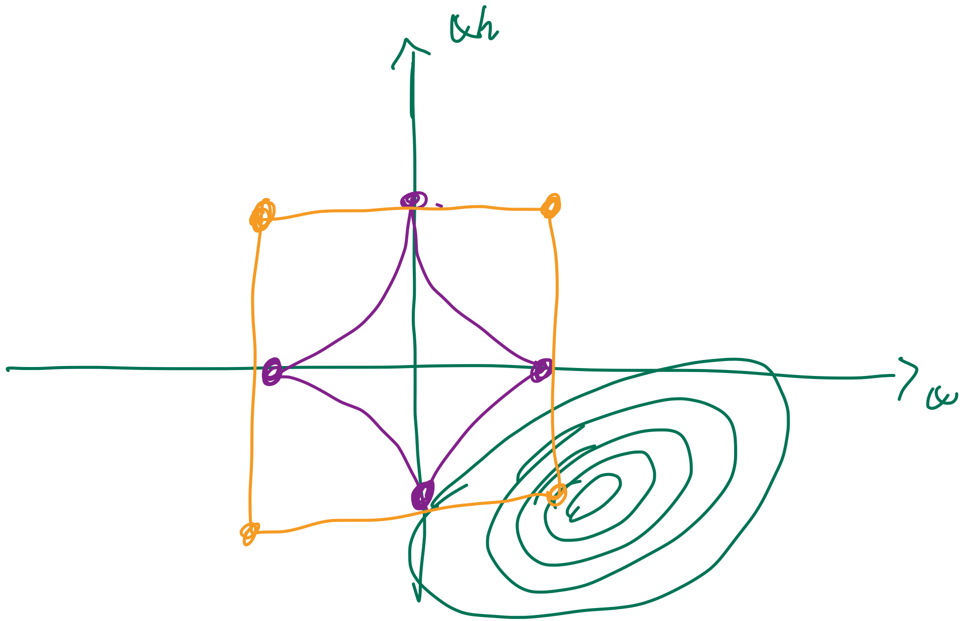
$$c_1^2 + c_2^2 = 1$$

$$|c_1| + |c_2| = 1$$



$$\| \ \|_\infty = \max_i (x_i)$$

$$\min_{\underline{w}} \; \| \qquad \|^2 + \lambda \, \|\underline{w}\|$$



$D_n \quad \{(\underline{x}_1, y_1), \ldots, (x_n, y_n)\}$ i.i.d.

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

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

$C(f) = \ldots$

$$\min_{f} \quad \hat{R}(f) + \lambda C(f)$$

i.i.d.

D_n

70% | 15% | 15%

LEARNING | VALIDATION | TEST

SVM ($\lambda = 3$, RBF, $\gamma = 4$) | 7% err

K.R. ($\lambda = 4$, RBF, $\gamma = 3$) | 3% –

2% .. → 2%

RUN

MODEL SELECTION & ERROR ESTIMATION

WISDOM OF THE CROWDS

$D_1 = 1\%$

$D_2 = 1\%$

$D_3 = 0,1\%$

$\simeq$ INDEP.

D1	D2	D3	DE	
0	0	0	0	
0	0	1	0	
0	1	0	0	
0	1	1	1	$\to 0,99 \cdot 0,01 \cdot 0,001 +$
1	0	0	0	
1	0	1	1	$\to 0,01 \cdot 0.99 \cdot 0,001 +$
1	1	0	1	$\to 0,01 \cdot 0,01 \cdot 0,999 +$
1	1	1	1	$\to 0,01 \cdot 0,01 \cdot 0,001 \simeq 0,0001 \ll 0,001$

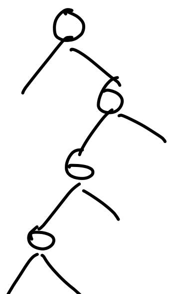
$$f(\underline{x}) = \underline{w}\,\underline{x}$$

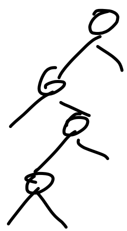
$$= \sum_{i=1}^{d} w_i\, x_i$$

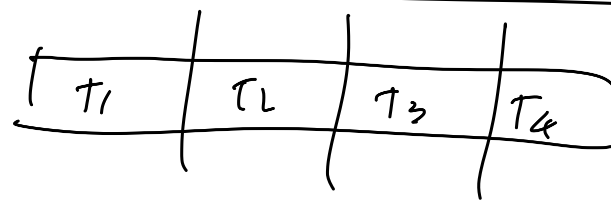
$$=$$

$\|\underline{w}\|_0$, $\|\underline{w}\|_2$

$$\underline{w}\ \underline{\phi}(x)$$



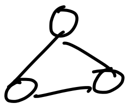


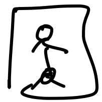


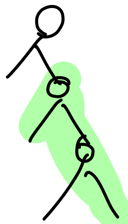
R.F.

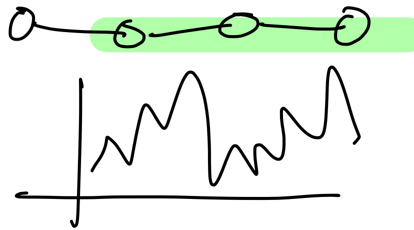
$D_n = \{(\underline{x}_1$

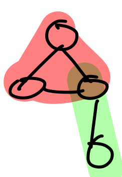
b



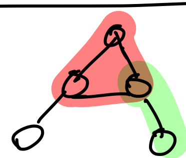








X_1



X_2

$\binom{n}{p} = n^p$

$X \longrightarrow \phi(x) \in \mathbb{R}$

$D_n = \{ (\bar{X}_1, y_1) \dots (\bar{X}_n, y_n) \}$ $\quad \bar{X} \in \mathbb{R}^d$ $\quad y <$ regr / class

88

$\square$

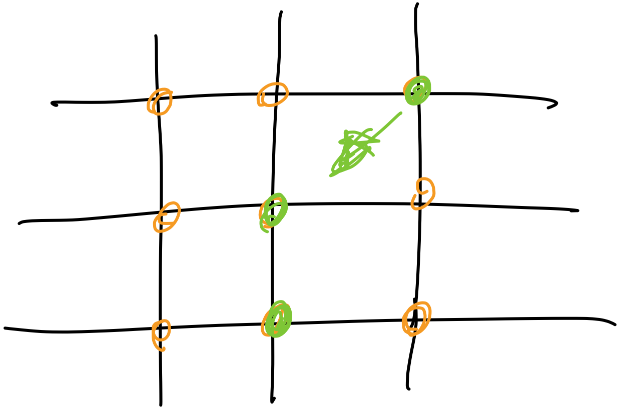
$f(x) = \dots$

$\ell(f(x), y) = \dots$

$C(f) = \dots$

$\min_f \hat{R}(f) + \lambda C(f)$

R. L.



$(p \cdot m_p)$

Sobsovet

1060