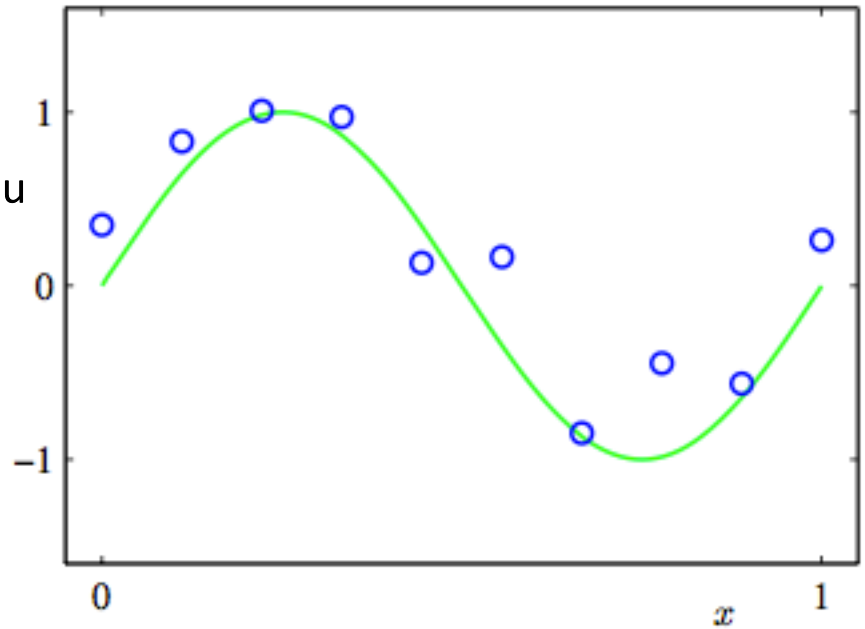
Gasirea polinomului optim



$$\mathcal{S} = \{(x_1, u_1), \ldots, (x_m, u_m)\}$$

multime cu exemple de antrenare, m = 10

$$u_i = f(x_i) + \epsilon_i$$

functia tinta
(vrem sa o invatam)

zgomot aleator
(corupe datele)

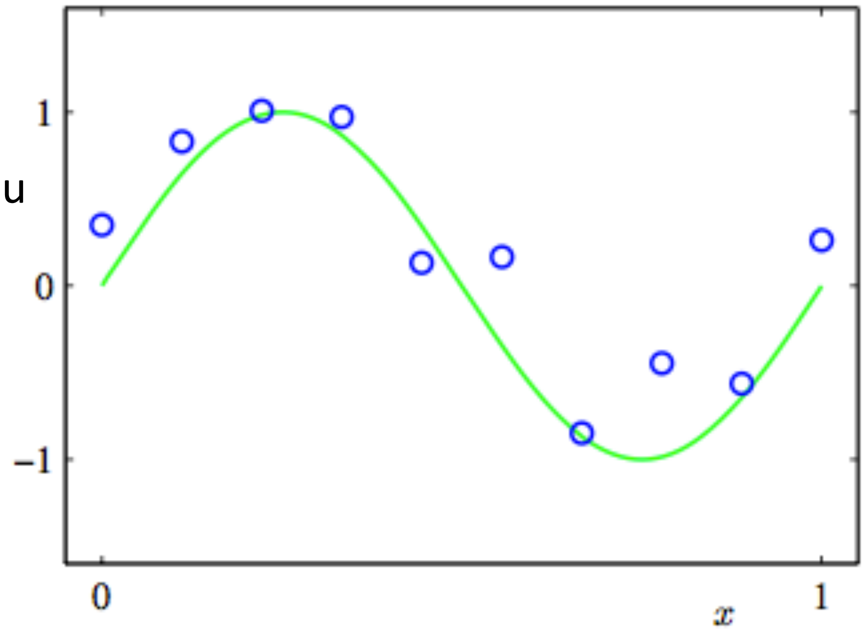
$$\epsilon_i \sim \mathcal{N}(\mu, \sigma)$$

Problema de regresie:
pe baza lui S gaseste (invata) functia h care stabileste corespondenta

$$u = h(x)$$

- foloseste h pentru predictie pentu noi valori ale lui x

Functia tinta



$$\mathcal{S} = \{(x_1, u_1), \ldots, (x_m, u_m)\}$$

multime cu exemple de antrenare, m = 10

$$u_i = f(x_i) + \epsilon_i$$

functia tinta
(vrem sa o invatam)

zgomot aleator
(corupe datele)

$$\epsilon_i \sim \mathcal{N}(\mu, \sigma)$$

$$f(x) = sin(2\pi x)$$

$$sin(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{(2n+1)!} x^n = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \ldots$$

$$sin(2\pi x) \approx 6.28x - 41.34x^3 + 81.60x^5 - \ldots$$

Spatiul de functii

$$\mathcal{H}_i = \{h(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots w_i x^i\}$$

Spatiul functiilor polinomiale (curbe) de grad i
w e parametru, h e liniara in w, h e neliniara in x

$$\mathcal{H}_0 \subseteq \mathcal{H}_1 \subseteq \mathcal{H}_2 \subseteq \ldots$$

functie constanta | drepte | parabole

Riscul empiric al unei ipoteze h:

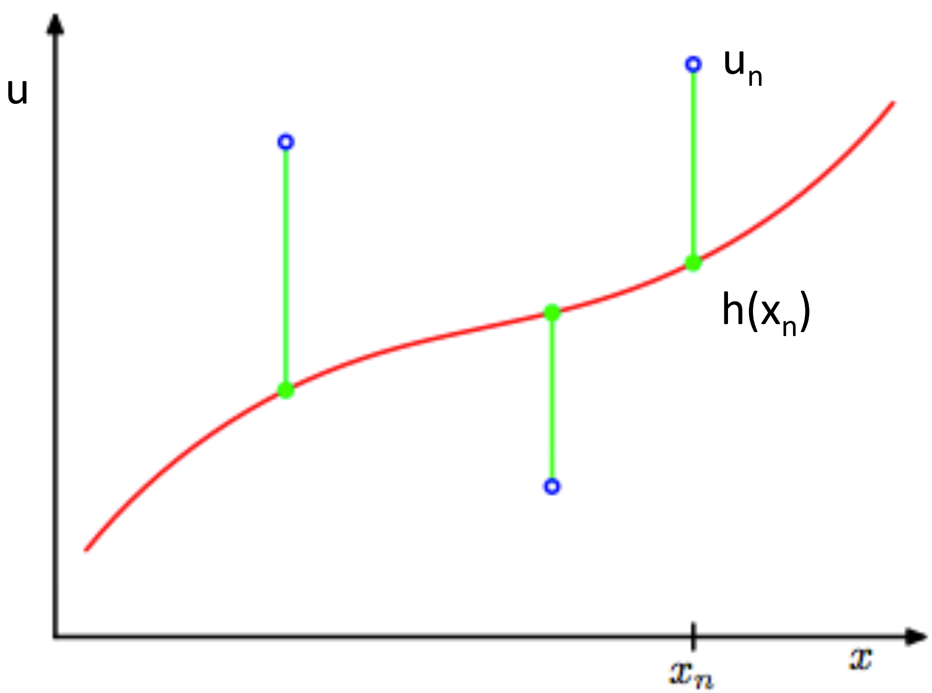
$$R_{emp}(h) = \frac{1}{|\mathcal{S}|} \sum_{i=1}^{|\mathcal{S}|} l(u_i, h(x_i, \mathbf{w}))$$

functie cost (loss) – masoara costul pe care il implica luarea deciziei $h(x_i)$ in loc de u_i

Functia cost

$$R_{emp}(h) = \frac{1}{|\mathcal{S}|} \sum_{i=1}^{|\mathcal{S}|} l(u_i, h(x_i, \mathbf{w}))$$

funcție cost (loss) – masoara costul pe care il implica luarea deciziei $h(x_i)$ in loc de u_i



Functia cost

$$R_{emp}(h) = \frac{1}{|\mathcal{S}|} \sum_{i=1}^{|\mathcal{S}|} l(u_i, h(x_i, \mathbf{w}))$$

functie cost (loss) – masoara costul pe care il implica luarea deciziei $h(x_i)$ in loc de u_i

Exemple de functii cost:

$$l(u_i, h(x_i, \mathbf{w})) = \sum_{i=1}^{|\mathcal{S}|} (u_i - h(x_i, \mathbf{w}))^2$$

$$l(u_i, h(x_i, \mathbf{w})) = \sum_{i=1}^{|\mathcal{S}|} |u_i - h(x_i, \mathbf{w})|$$

Principiul ERM

- gaseste ipoteza h* care minimizeaza riscul empiric (eroarea de antrenare)

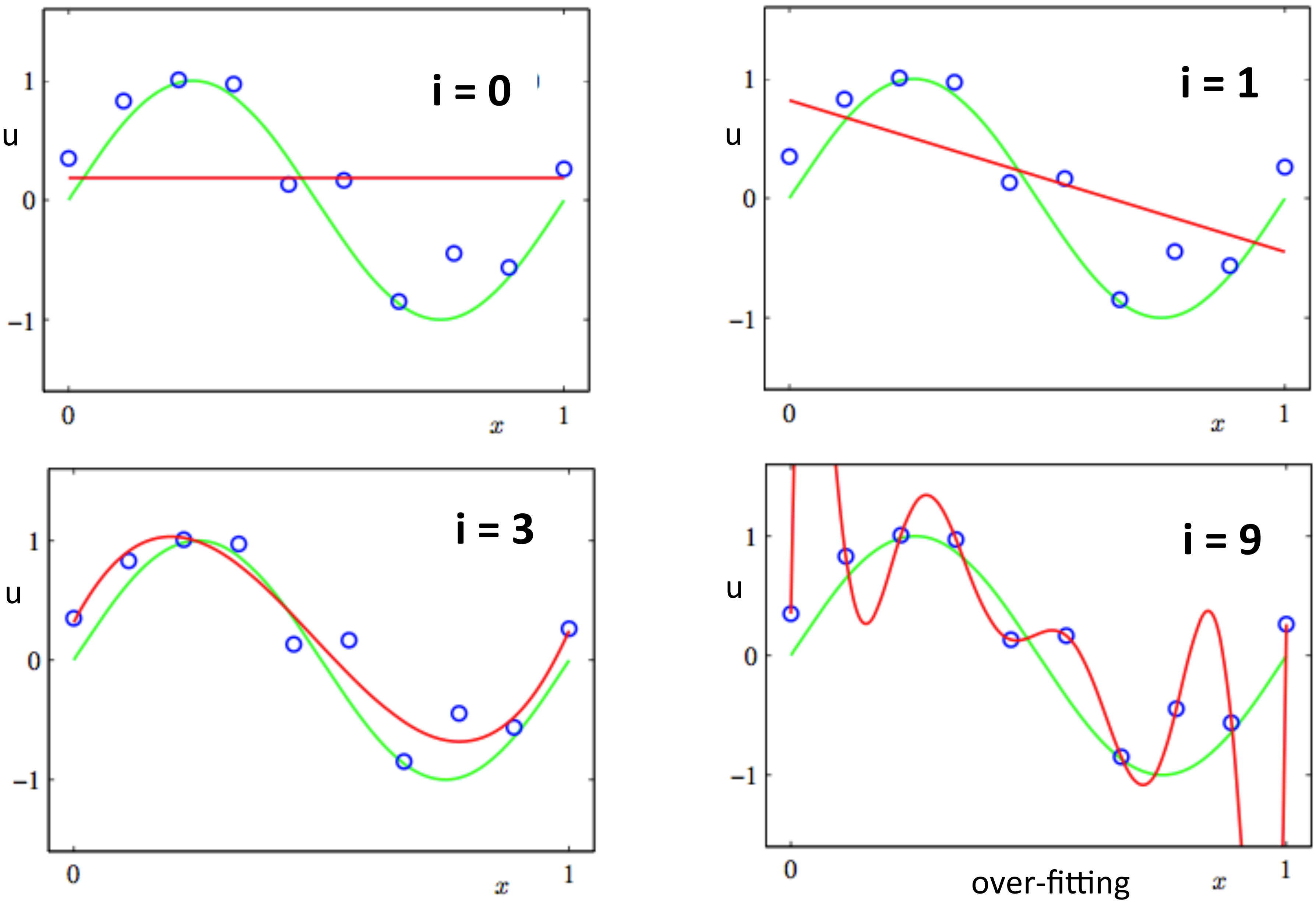
$$h^*_{\mathcal{S},\mathcal{H}} = arg \min_{h \in \mathcal{H}} R_{emp}(h)$$

$$R_{emp}(h) = \frac{1}{|\mathcal{S}|} \sum_{i=1}^{|\mathcal{S}|} l(u_i, h(x_i, \mathbf{w}))$$

- afla parametri w care minimizeaza riscul empiric folosind functia de cost

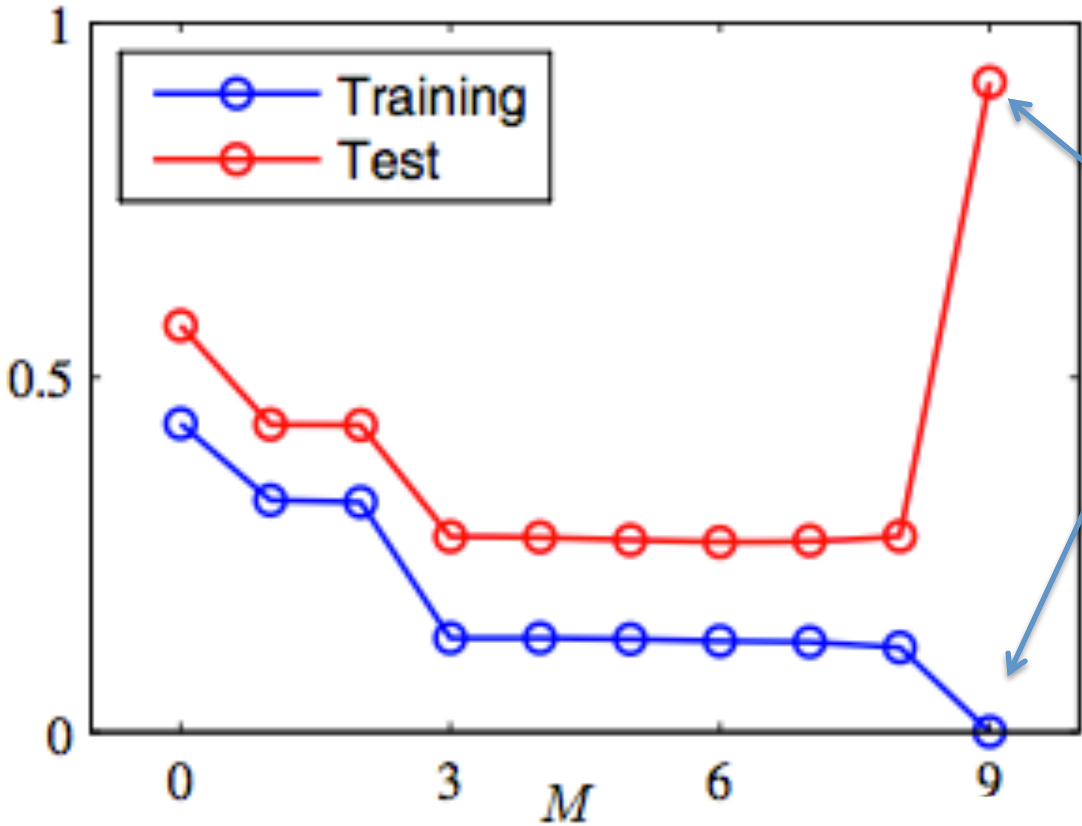
$$l(u_i, h(x_i, \mathbf{w})) = \sum_{i=1}^{|\mathcal{S}|} (u_i - h(x_i, \mathbf{w}))^2$$

Alegerea modelului $\mathcal{H}_i$ si aflarea lui $h^*_{\mathcal{S},\mathcal{H}_i}$



Evolutia riscului empiric

Evaluam ipotezele pe o multime de test de 100 de exemple



i

Over-fitting:
- eroare pe multimea de antrenare mica
- eroare pe multimea de testare mare

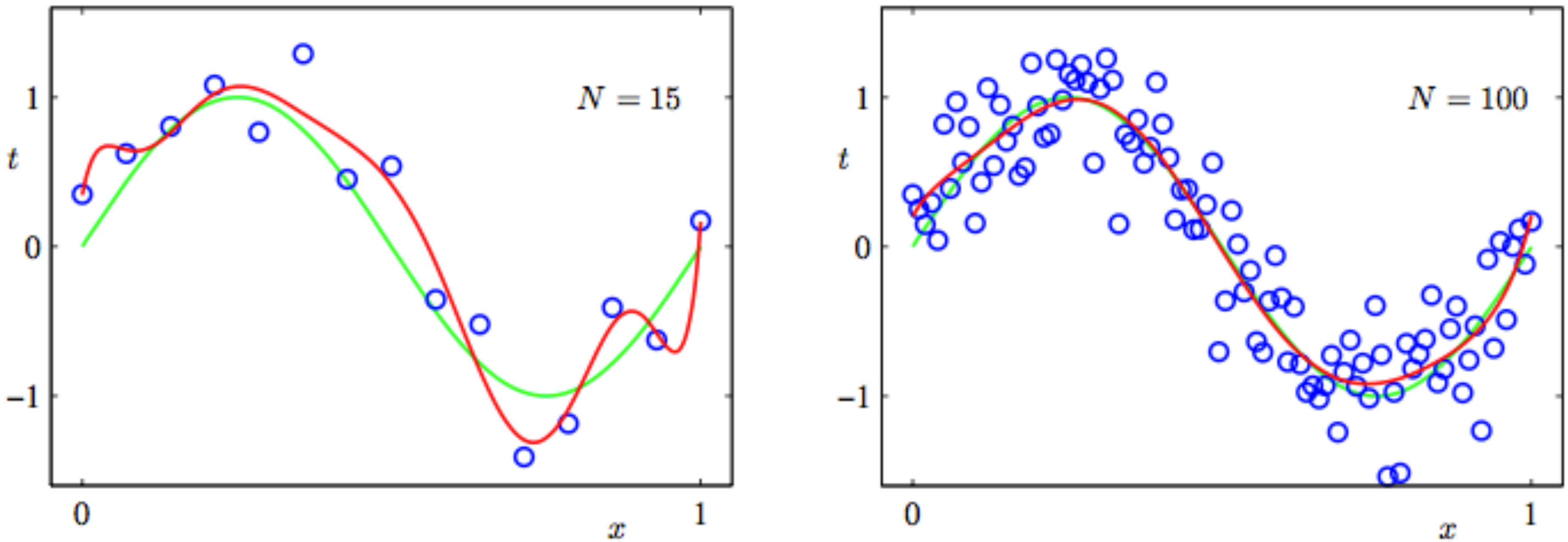
Coeficientii lui $h^*_{\mathcal{S},\mathcal{H}_i}$

$$\mathcal{H}_i = \{h(x, \mathbf{w}) = w_0 + w_1 x + w_2 x^2 + \ldots w_i x^i\}$$

	i = 0	i = 1	i = 6	i = 9
$w_0^\star$	0.19	0.82	0.31	0.35
$w_1^\star$		-1.27	7.99	232.37
$w_2^\star$			-25.43	-5321.83
$w_3^\star$			17.37	48568.31
$w_4^\star$				-231639.30
$w_5^\star$				640042.26
$w_6^\star$				-1061800.52
$w_7^\star$				1042400.18
$w_8^\star$				-557682.99
$w_9^\star$				125201.43

$$sin(2\pi x) \approx 6.28x - 41.34x^3 + 81.60x^5 - \ldots$$

Comportamentul unui model in functie de marimea lui S



Problema de over-fitting se elimina treptat pe masura ce creste numarul de exemple de antrenare

Metode de regularizare

$$l(u_i, h(x_i, \mathbf{w})) = \sum_{i=1}^{|\mathcal{S}|} (u_i - h(x_i, \mathbf{w}))^2 + \underbrace{\lambda ||w||^2}_{\text{penalitate}}$$

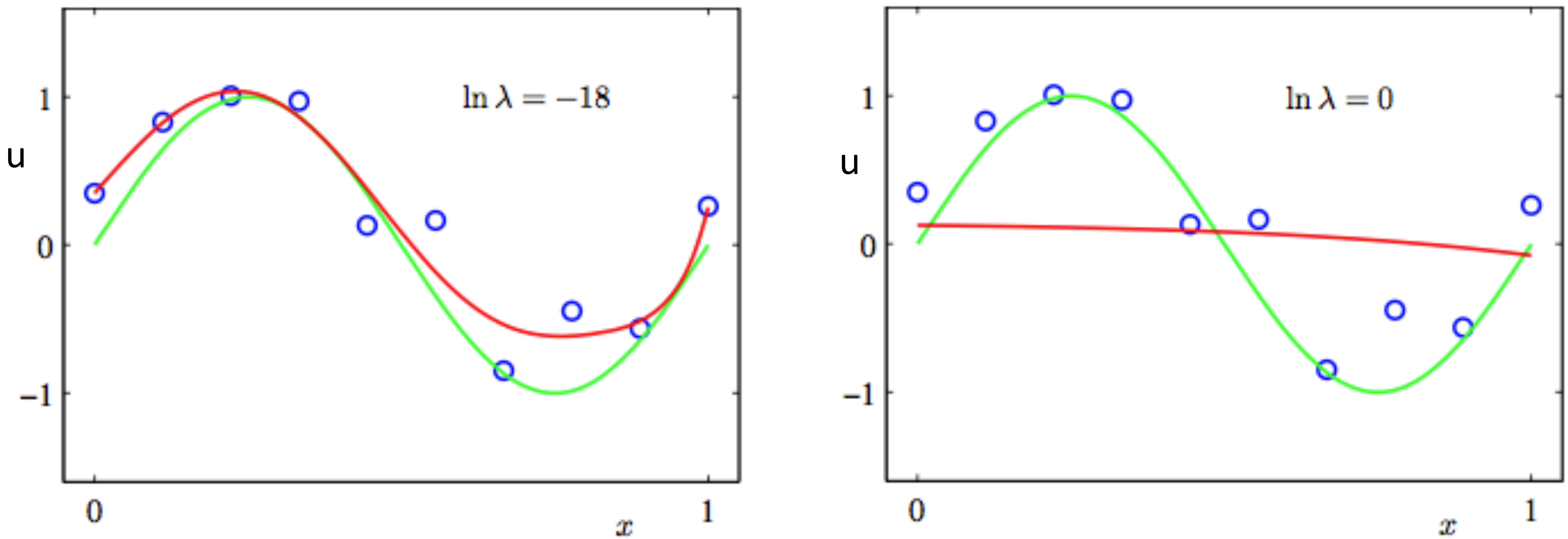
$$||w||^2 = w_0^2 + w_1^2 + \ldots + w_i^2$$

λ - controleaza importanta termenului de regularizare/penalitate

Impactul includerii unei termen de regularizare

$$l(u_i, h(x_i, \mathbf{w})) = \sum_{i=1}^{|\mathcal{S}|}(u_i - h(x_i, \mathbf{w}))^2 + \underbrace{\lambda||w||^2}_{\text{penalitate}}$$

$h^*_{\mathcal{S},\mathcal{H}_9}$ $h^*_{\mathcal{S},\mathcal{H}_9}$



Impactul includerii unei termen de regularizare

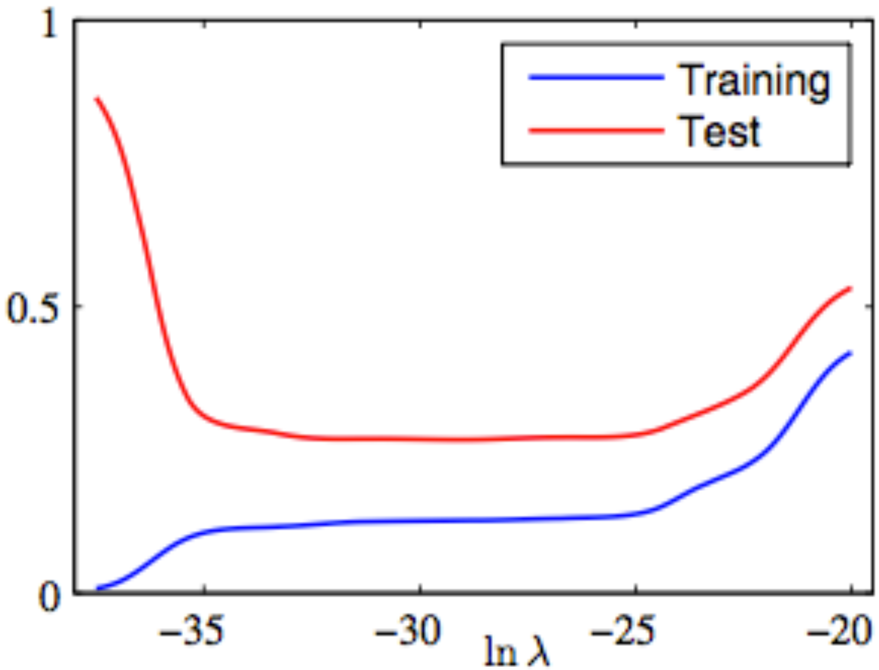
	i = 0	i = 1	i = 6	i = 9
$w_0^\star$	0.19	0.82	0.31	0.35
$w_1^\star$		-1.27	7.99	232.37
$w_2^\star$			-25.43	-5321.83
$w_3^\star$			17.37	48568.31
$w_4^\star$				-231639.30
$w_5^\star$				640042.26
$w_6^\star$				-1061800.52
$w_7^\star$				1042400.18
$w_8^\star$				-557682.99
$w_9^\star$				125201.43

i = 9

	$\ln \lambda = -\infty$	$\ln \lambda = -18$	$\ln \lambda = 0$
$w_0^\star$	0.35	0.35	0.13
$w_1^\star$	232.37	4.74	-0.05
$w_2^\star$	-5321.83	-0.77	-0.06
$w_3^\star$	48568.31	-31.97	-0.05
$w_4^\star$	-231639.30	-3.89	-0.03
$w_5^\star$	640042.26	55.28	-0.02
$w_6^\star$	-1061800.52	41.32	-0.01
$w_7^\star$	1042400.18	-45.95	-0.00
$w_8^\star$	-557682.99	-91.53	0.00
$w_9^\star$	125201.43	72.68	0.01

$$sin(2\pi x) \approx 6.28x - 41.34x^3 + 81.60x^5 - \dots$$

Impactul includerii unei termen de regularizare



i = 9

	$\ln\lambda = -\infty$	$\ln\lambda = -18$	$\ln\lambda = 0$
$w_0^\star$	0.35	0.35	0.13
$w_1^\star$	232.37	4.74	-0.05
$w_2^\star$	-5321.83	-0.77	-0.06
$w_3^\star$	48568.31	-31.97	-0.05
$w_4^\star$	-231639.30	-3.89	-0.03
$w_5^\star$	640042.26	55.28	-0.02
$w_6^\star$	-1061800.52	41.32	-0.01
$w_7^\star$	1042400.18	-45.95	-0.00
$w_8^\star$	-557682.99	-91.53	0.00
$w_9^\star$	125201.43	72.68	0.01

Alegerea modelului

Impartim datele initiale in 2 multimi: multimea de antrenare si multimea de validare.

Alegem i sau λ pe baza erorii pe multimii de validare