Least Squares becslés

A négyzetes hibafüggvény:

$$\frac{1}{2}\sum_{i=1}^{P}\{\, d_i - \mathbf{w}^{\mathrm{T}}\, \boldsymbol{\varphi}(\mathbf{x}_i)\}^2$$

A négyzetes hibafüggvény mellett a minimumot biztosító megoldás

$$\mathbf{W}=\left(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\right)^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{d}$$

LS becslés

A gradiens számítása és nullává tétele eredményeképp

A megoldás

$$\mathbf{w}_{\mathrm{LS}}^{*} = \left(\mathbf{\Phi}^{T}\mathbf{\Phi}\right)^{-1}\mathbf{\Phi}^{T}\mathbf{d}$$

A Moore-Penrose pseudo-inverse, $\mathbf{\Phi}^{\dagger}$.

Regularizált Least Squares

A hibafüggvény:

$$E_D(\mathbf{w}) + \lambda E_W(\mathbf{w})$$

Adatoktól függő + Regularizációs tag

A négyzetes hibafüggvény és a négyzetes regularizációs összefüggés mellett

$$\frac{\lambda}{2}\mathbf{w}^{\mathrm{T}}\mathbf{w} + \frac{1}{2}\sum_{i=1}^{P}\{\, d_i - \mathbf{w}^{\mathrm{T}}\, \boldsymbol{\varphi}(\mathbf{x}_i)\}^2$$

λ: regularizációs együttható

Ennek minimumát biztosító megoldás

$$\mathbf{w} = \left(\lambda\mathbf{I} + \boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\right)^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{t}.$$

Maximum Likelihood becslés

Feltesszük, hogy a megfigyelések egy determinisztikus függvényből származnak Gauss additív zajjal terhelve

$$d = y(\mathbf{x}, \mathbf{w}) + n \qquad p(n|\beta) = N(n|0, \beta^{-1})$$

Amit írhatunk így is:

$$p(d|\mathbf{x}, \mathbf{w}, \beta) = N(d|y(\mathbf{x}, \mathbf{w}), \beta^{-1})$$

Adott $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_P]^T$ bemenetek és $\mathbf{d} = [d_1, d_2, \ldots, d_P]^T$ kívánt válasz megfigyeléseknél a likelihood függvény a következő

$$p(\mathbf{d}|\mathbf{X}, \mathbf{w}, \beta) = \prod_i N(d_i|y(\mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i), \beta^{-1})$$

Maximum likelihood becslés

Vegyük a logaritmusát:

$$\ln p(\mathbf{d}|\mathbf{X},\mathbf{w},\beta) = \sum_i \ln N(d_i | y(\mathbf{w}^T \boldsymbol{\varphi}(\mathbf{x}_i), \beta^{-1})$$

$$= \frac{P}{2}\ln\beta - \frac{P}{2}\ln(2\pi) - \frac{1}{2}\beta\sum_i (d_i - \mathbf{w}^T\boldsymbol{\varphi}(\mathbf{x}_i))^2$$

Vagyis az egyes megfigyeléseknél az eltérések négyzetének összegét nézzük

ML becslés

A gradiens számítása és nullává tétele eredményeképp

A megoldás

$$\mathbf{w}_{\mathrm{ML}}^{*} = \left(\boldsymbol{\Phi}^{T}\boldsymbol{\Phi}\right)^{-1}\boldsymbol{\Phi}^{T}\mathbf{d}$$

A Moore-Penrose pseudo-inverse, $\boldsymbol{\Phi}^{\dagger}$.

Ahol

$$\mathbf{w}_{\mathrm{ML}}^{*} = \boldsymbol{\Phi}^{\dagger}\mathbf{d} = (\boldsymbol{\Phi}^{T}\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^{T}\mathbf{d} = \mathbf{w}_{\mathrm{LS}}^{*}$$

Maximum likelihood becslés

β is meghatározható ha e szerint keressük a szélsőértéket:

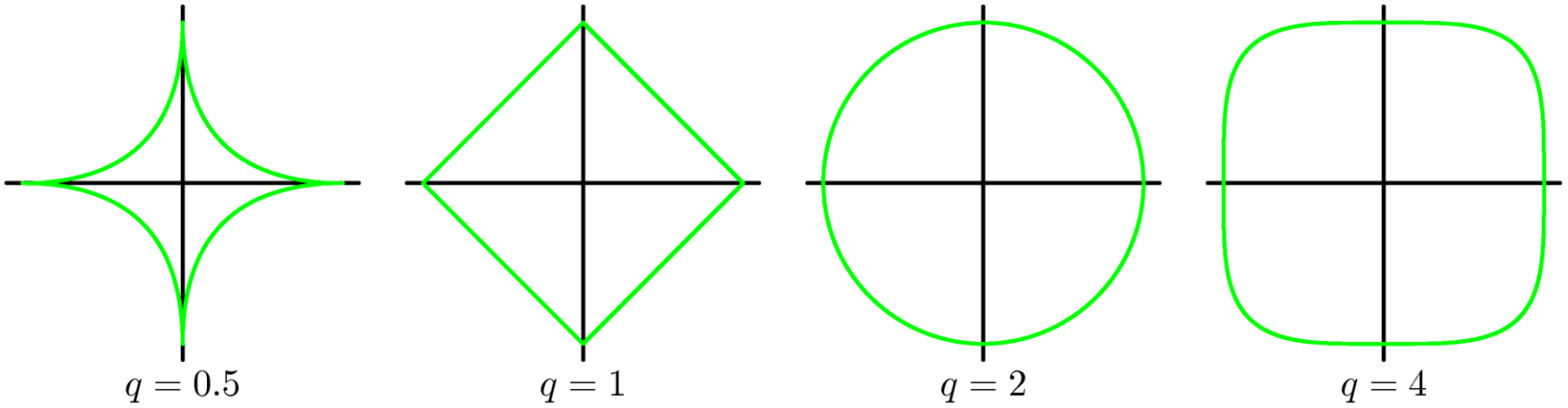
$$\frac{1}{\beta_{\mathrm{ML}}} = \frac{1}{P}\sum_{i=1}^{P}\{d_i - \mathbf{w}^{\mathrm{T}}\,\boldsymbol{\varphi}(\mathbf{x}_i)\}^2$$

Ami a jól ismert tapasztalati szórásnégyzet összefüggés

Regularizált Least Squares

Egyéb szokásos regularizációs lehetőségek

$$\frac{\lambda}{2}\sum_{j=1}^{M}|w_j|^q + \frac{1}{2}\sum_{i=1}^{P}\{d_i - \mathbf{w}^{\mathrm{T}}\,\boldsymbol{\varphi}(\mathbf{x}_i)\}^2$$

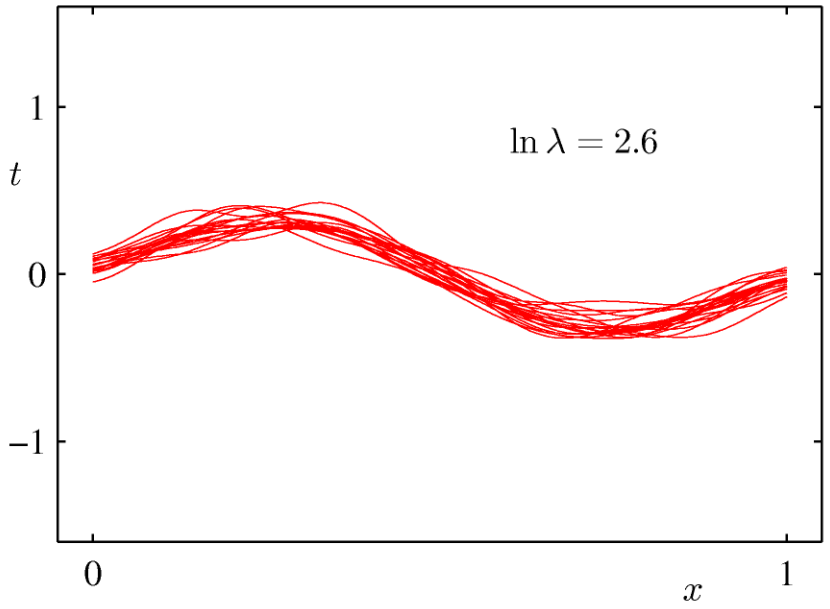


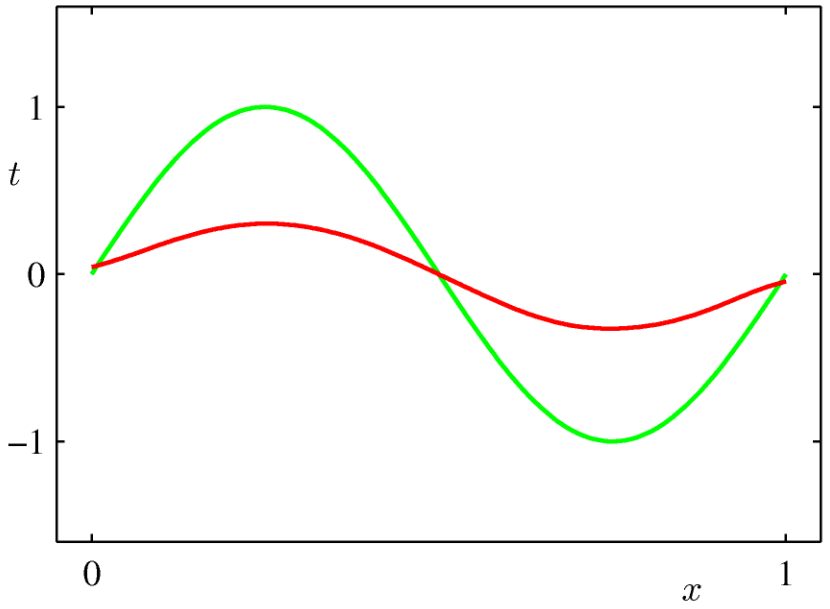
Lasso

Négyzetes

A tozítás-variancia dekompozíció

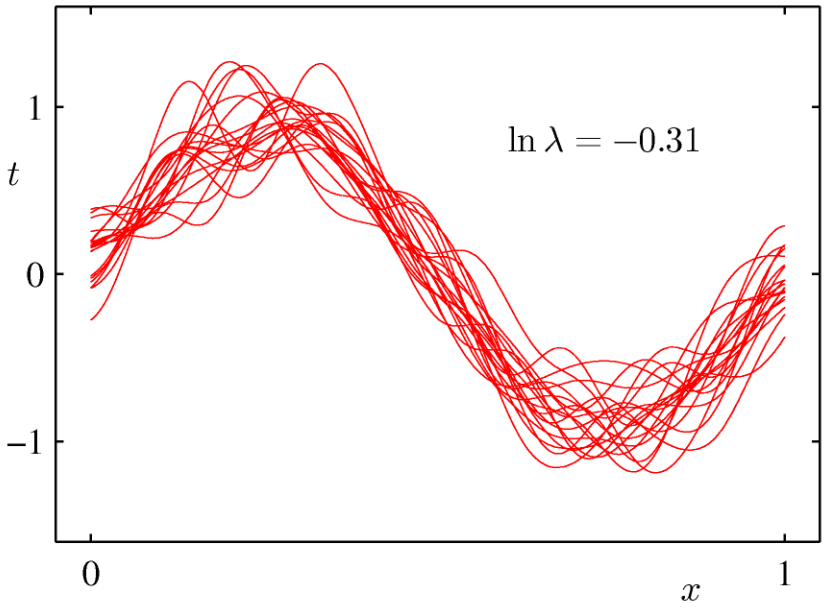
Egy példa: 25 adatkészlet, a regularizációs együttható függvényében.

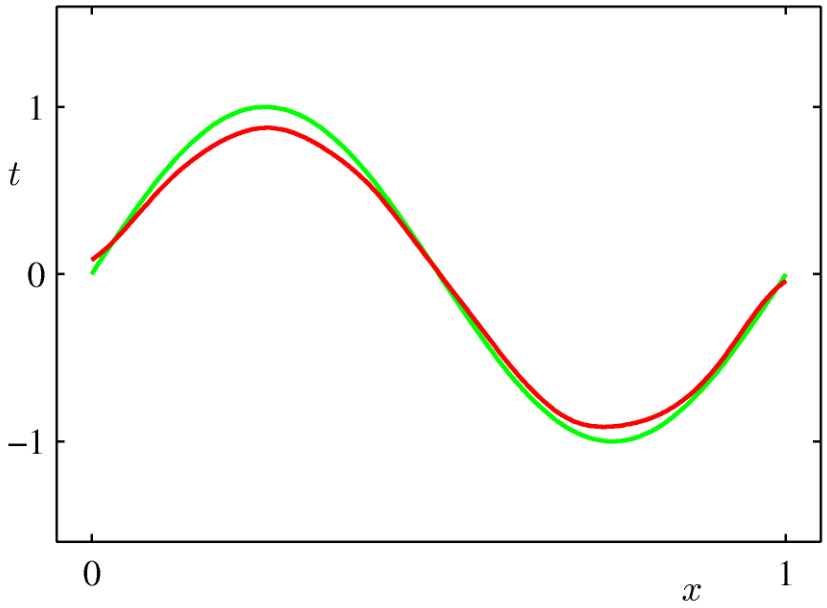




A tozítás-variancia dekompozíció

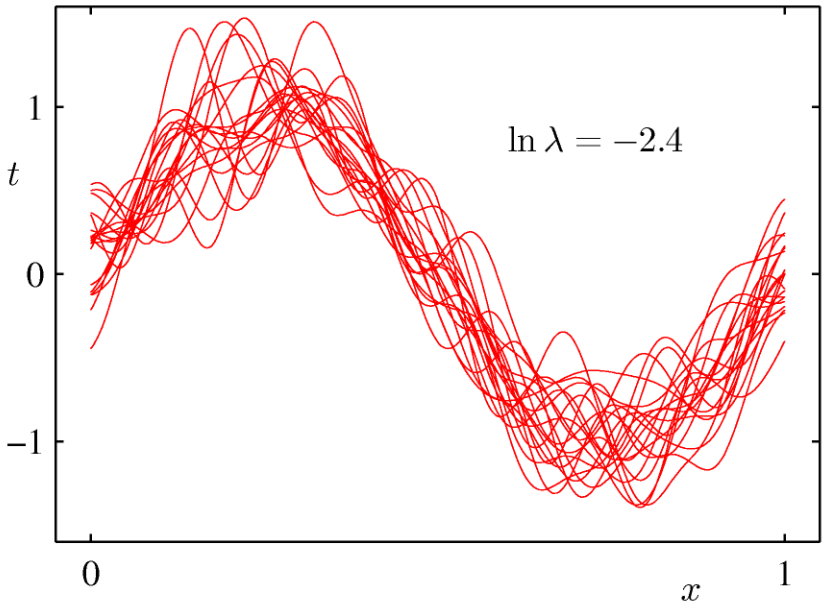
Egy példa: 25 adatkészlet, a regularizációs együttható függvényében

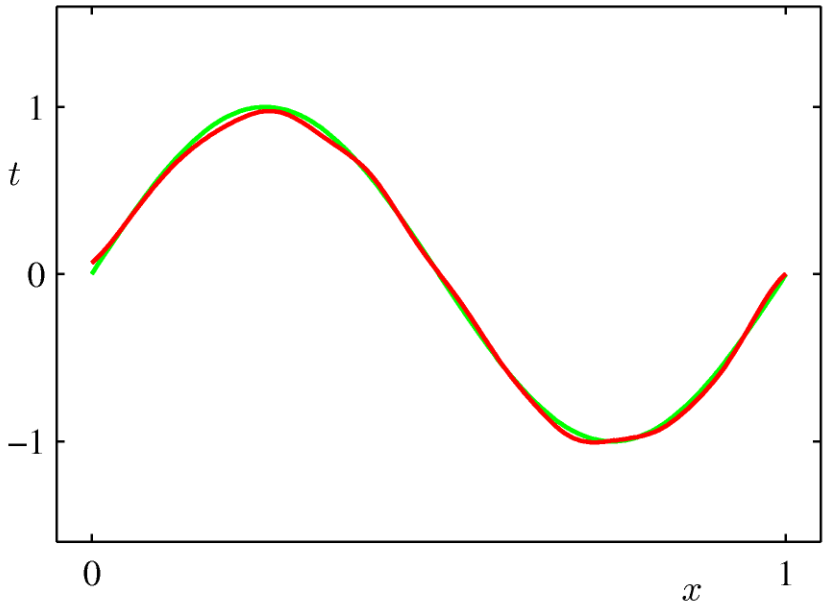




A tozítás-variancia dekompozíció

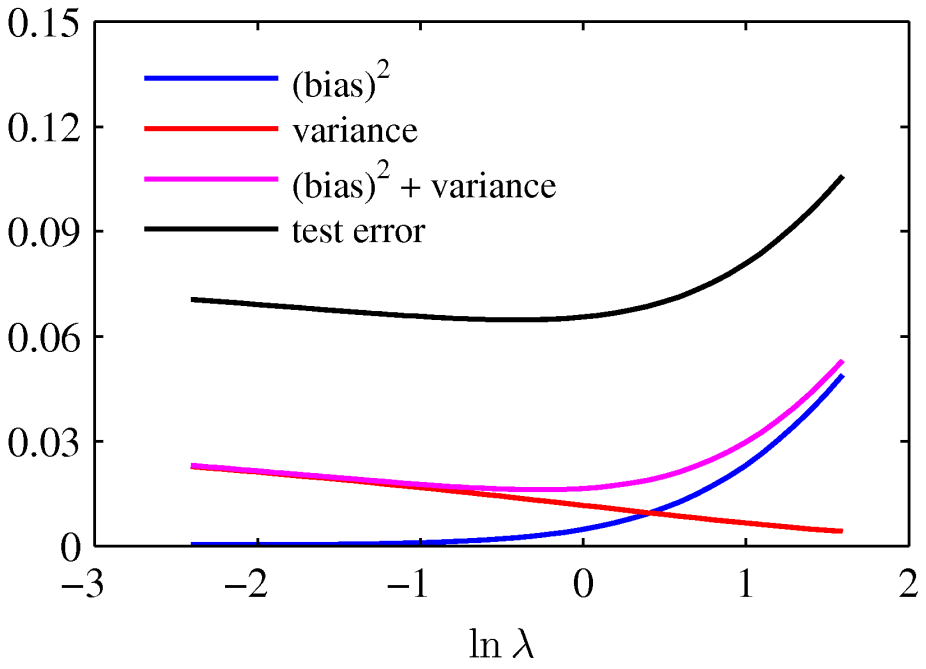
Egy példa: 25 adatkészlet, a regularizációs együttható függvényében





A tozítás-variancia dekompozíció

Egy túlregularizált modell (nagy λ) erősen torzított lesz, míg egy alulregularizált modell (kis λ) nagy variancivával fog rendelkezni.



Bayes lineáris regresszió

A Bayes megközelítésnél feltételezzük, hogy **w** valószínűségi változó, ismert a priori sűrűségfüggvénnyel

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\mathbf{m}_0, \mathbf{S}_0).$$

A megfigyelések birtokában meghatározható az a poszteriori sűrűségfüggvény. Ehhez a Bayes tételt alkalmazzuk:

$$p(Y|X) = \frac{p(X|Y)p(Y)}{p(X)} \qquad p(X) = \sum_Y p(X|Y)p(Y)$$

posterior $\propto$ likelihood × prior $\qquad p(\mathbf{w}|d) = \frac{p(d|\mathbf{w})p(\mathbf{w})}{p(d)} = \frac{p(d|\mathbf{w})p(\mathbf{w})}{\int p(d|\mathbf{w})p(\mathbf{w})d\mathbf{w}}$

Mivel itt minden Gauss a posterior is Gauss lesz

$$p(\mathbf{w}|\mathbf{d}) = N(\mathbf{w}|\mathbf{m}_P, \mathbf{S}_P)$$

ahol

$$\begin{aligned} \mathbf{m}_P &= \mathbf{S}_P \left(\mathbf{S}_0^{-1}\mathbf{m}_0 + \beta\mathbf{\Phi}^{\mathrm{T}}\mathbf{d}\right) \\ \mathbf{S}_P^{-1} &= \mathbf{S}_0^{-1} + \beta\mathbf{\Phi}^{\mathrm{T}}\mathbf{\Phi}. \end{aligned}$$

Bayes lineáris regresszió

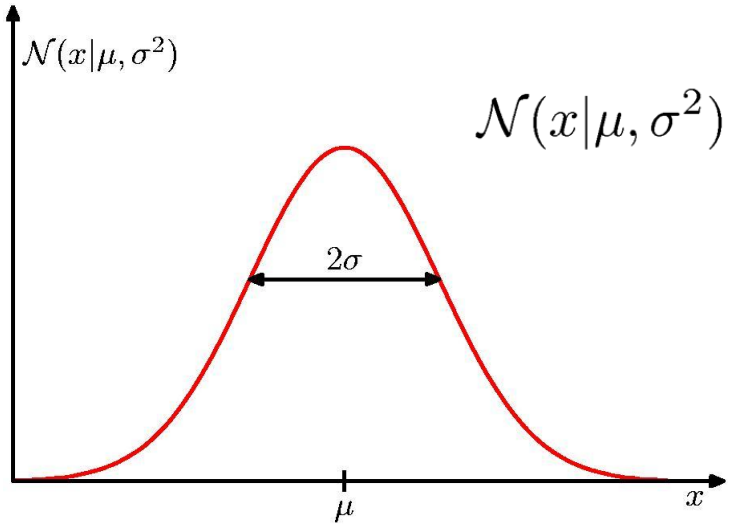
Gyakori választás a priorra

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\mathbf{0}, \alpha^{-1}\mathbf{I})$$

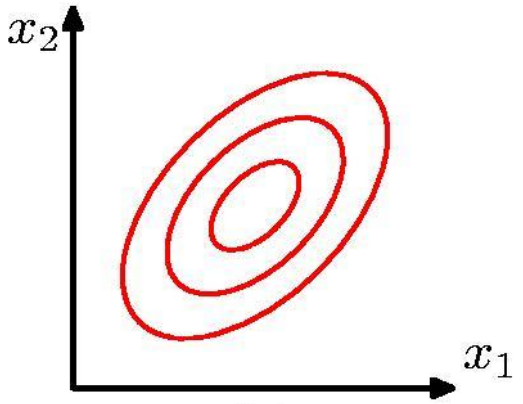
Ekkor

$$\begin{aligned} \mathbf{m}_P &= \beta \mathbf{S}_P \mathbf{\Phi}^{\mathrm{T}} \mathbf{t} \\ \mathbf{S}_P^{-1} &= \alpha \mathbf{I} + \beta \mathbf{\Phi}^{\mathrm{T}} \mathbf{\Phi}. \end{aligned}$$

Gauss eloszlás



$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\}$$



$$\mathcal{N}(\mathbf{x}|\boldsymbol{\mu},\boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{D/2}} \frac{1}{|\boldsymbol{\Sigma}|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

Partitionált Gauss eloszlás

$$p(\mathbf{x}) = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{pmatrix} \qquad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_a \\ \boldsymbol{\mu}_b \end{pmatrix} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{aa} & \boldsymbol{\Sigma}_{ab} \\ \boldsymbol{\Sigma}_{ba} & \boldsymbol{\Sigma}_{bb} \end{pmatrix}$$

$$\boldsymbol{\Lambda} \equiv \boldsymbol{\Sigma}^{-1} \qquad \boldsymbol{\Lambda} = \begin{pmatrix} \boldsymbol{\Lambda}_{aa} & \boldsymbol{\Lambda}_{ab} \\ \boldsymbol{\Lambda}_{ba} & \boldsymbol{\Lambda}_{bb} \end{pmatrix}$$

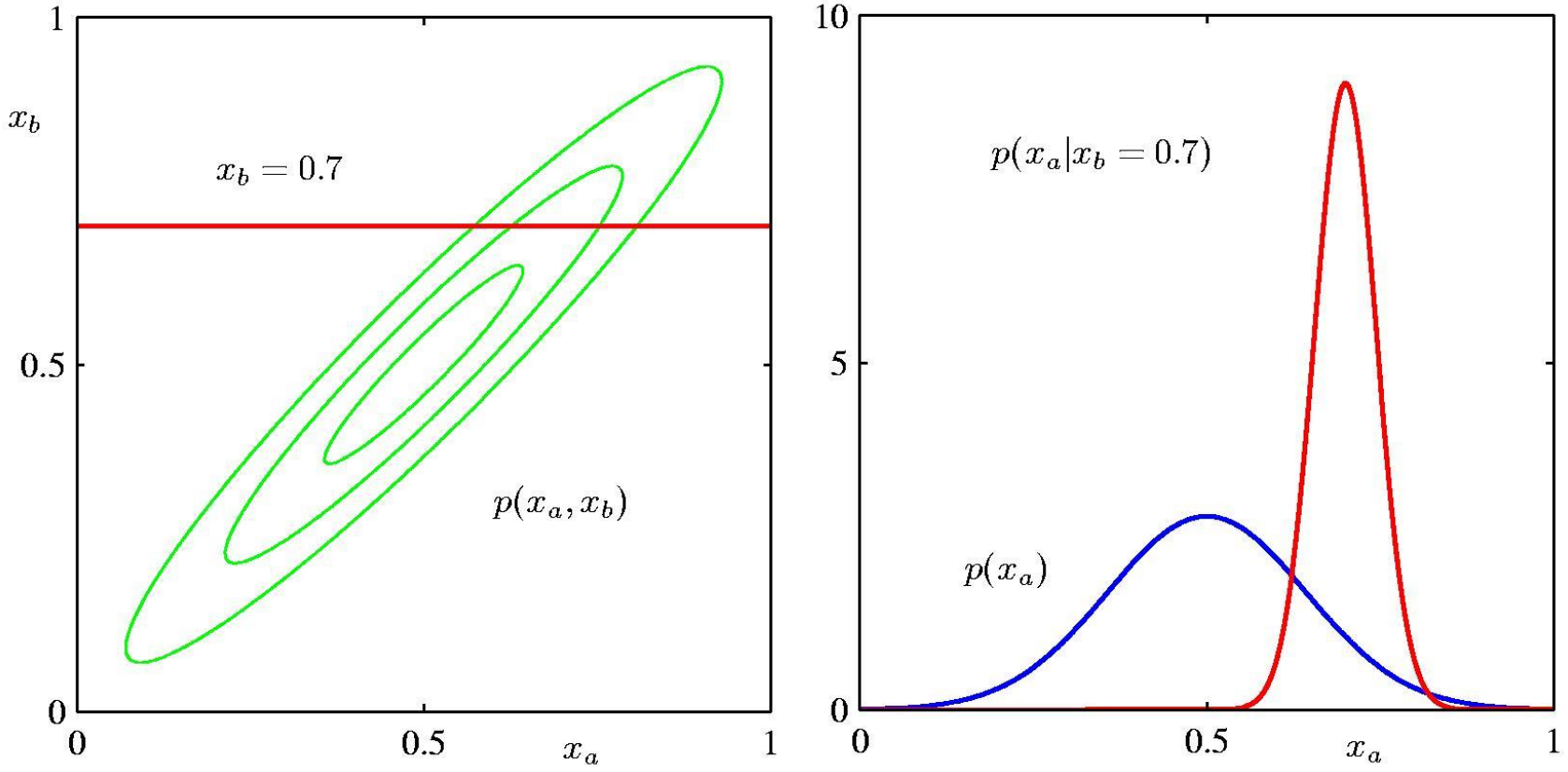
Partitionált feltételes eloszlás és marginális

$$p(\mathbf{x}_a|\mathbf{x}_b) = \mathcal{N}(\mathbf{x}_a|\boldsymbol{\mu}_{a|b}, \boldsymbol{\Sigma}_{a|b})$$

$$\begin{aligned}
\boldsymbol{\Sigma}_{a|b} &= \boldsymbol{\Lambda}_{aa}^{-1} = \boldsymbol{\Sigma}_{aa} - \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}\boldsymbol{\Sigma}_{ba} \\
\boldsymbol{\mu}_{a|b} &= \boldsymbol{\Sigma}_{a|b}\left\{\boldsymbol{\Lambda}_{aa}\boldsymbol{\mu}_a - \boldsymbol{\Lambda}_{ab}(\mathbf{x}_b - \boldsymbol{\mu}_b)\right\} \\
&= \boldsymbol{\mu}_a - \boldsymbol{\Lambda}_{aa}^{-1}\boldsymbol{\Lambda}_{ab}(\mathbf{x}_b - \boldsymbol{\mu}_b) \\
&= \boldsymbol{\mu}_a + \boldsymbol{\Sigma}_{ab}\boldsymbol{\Sigma}_{bb}^{-1}(\mathbf{x}_b - \boldsymbol{\mu}_b)
\end{aligned}$$

$$\begin{aligned}
p(\mathbf{x}_a) &= \int p(\mathbf{x}_a, \mathbf{x}_b)\,\mathrm{d}\mathbf{x}_b \\
&= \mathcal{N}(\mathbf{x}_a|\boldsymbol{\mu}_a, \boldsymbol{\Sigma}_{aa})
\end{aligned}$$

Partitionált feltételes eloszlás és marginális



Bayes tétel Gauss változókra

Adott

$$
\begin{aligned}
p(\mathbf{x}) &= \mathcal{N}\left(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Lambda}^{-1}\right) \\
p(\mathbf{y}|\mathbf{x}) &= \mathcal{N}\left(\mathbf{y}|\mathbf{A}\mathbf{x}+\mathbf{b}, \mathbf{L}^{-1}\right)
\end{aligned}
$$

ebből

$$
\begin{aligned}
p(\mathbf{y}) &= \mathcal{N}(\mathbf{y}|\mathbf{A}\boldsymbol{\mu}+\mathbf{b}, \mathbf{L}^{-1}+\mathbf{A}\boldsymbol{\Lambda}^{-1}\mathbf{A}^{\mathrm{T}}) \\
p(\mathbf{x}|\mathbf{y}) &= \mathcal{N}(\mathbf{x}|\boldsymbol{\Sigma}\{\mathbf{A}^{\mathrm{T}}\mathbf{L}(\mathbf{y}-\mathbf{b})+\boldsymbol{\Lambda}\boldsymbol{\mu}\}, \boldsymbol{\Sigma})
\end{aligned}
$$

ahol

$$
\boldsymbol{\Sigma} = (\boldsymbol{\Lambda}+\mathbf{A}^{\mathrm{T}}\mathbf{L}\mathbf{A})^{-1}
$$

Bayes következtetés Gauss esetben

Tételezzük fel, hogy σ^2 ismert. Adott i.i.d. adat $\mathbf{x} = \{x_1, \ldots, x_P\}$, a likelihood függvény μ -re

$$p(\mathbf{x}|\mu) = \prod_{n=1}^{P} p(x_n|\mu) = \frac{1}{(2\pi\sigma^2)^{P/2}} \exp\left\{-\frac{1}{2\sigma^2}\sum_{n=1}^{P}(x_n - \mu)^2\right\}.$$

ez is Gauss.

Bayes következtetés Gauss esetben

Gauss prior μfelett,

$$p(\mu) = \mathcal{N}\left(\mu|\mu_0, \sigma_0^2\right).$$

Ebből a posterior

$$p(\mu|\mathbf{x}) \propto p(\mathbf{x}|\mu)p(\mu).$$

Teljes négyzetté alakítással

$$p(\mu|\mathbf{x}) = \mathcal{N}\left(\mu|\mu_N, \sigma_P^2\right)$$

Bayes következtetés Gauss esetben

… ahol

$$\mu_P = \frac{\sigma^2}{P\sigma_0^2 + \sigma^2}\mu_0 + \frac{P\sigma_0^2}{P\sigma_0^2 + \sigma^2}\mu_{\mathrm{ML}}, \qquad \mu_{\mathrm{ML}} = \frac{1}{P}\sum_{n=1}^{P} x_n$$

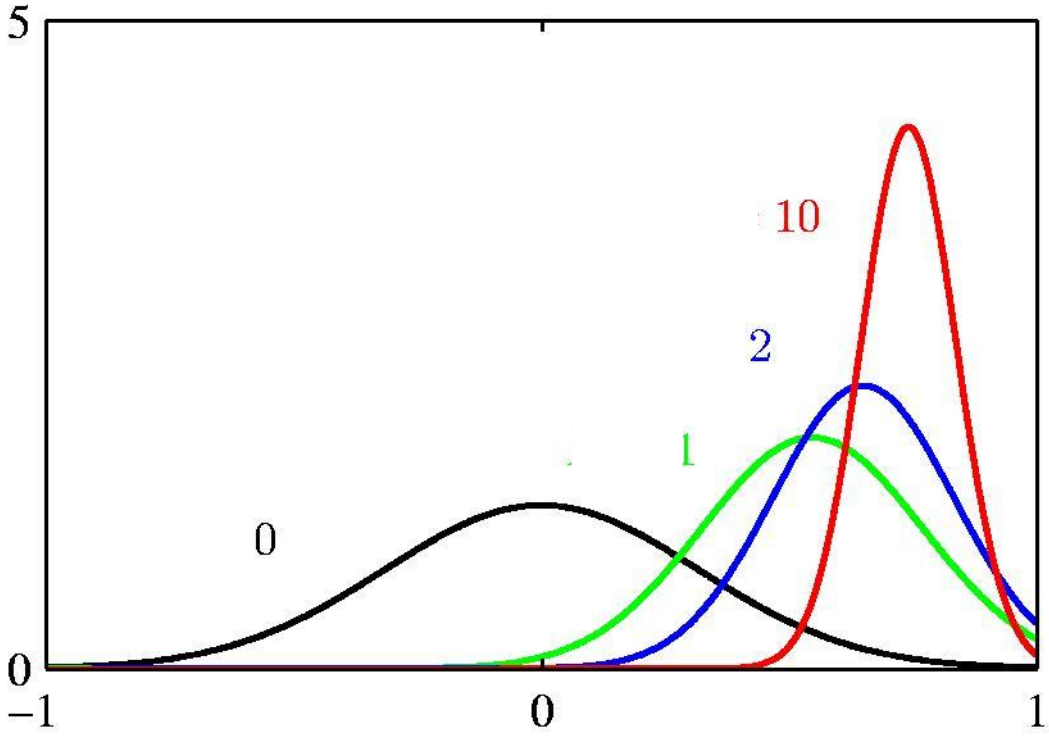
$$\frac{1}{\sigma_P^2} = \frac{1}{\sigma_0^2} + \frac{P}{\sigma^2}.$$

Megjegyzés:

	$P = 0$	$P \to \infty$
μ_P	μ_0	μ_{ML}
σ_P^2	σ_0^2	0

Bayes következtetés Gauss esetben

Példa: $p(\mu|\mathbf{x}) = \mathcal{N}\left(\mu|\mu_P, \sigma_P^2\right)$ for P = 0, 1, 2 and 10.



Bayes lineáris regresszió

Gyakori választás a priorra

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\mathbf{0}, \alpha^{-1}\mathbf{I})$$

Ekkor

$$\begin{aligned} \mathbf{m}_P &= \beta \mathbf{S}_P \mathbf{\Phi}^{\mathrm{T}} \mathbf{t} \\ \mathbf{S}_P^{-1} &= \alpha \mathbf{I} + \beta \mathbf{\Phi}^{\mathrm{T}} \mathbf{\Phi}. \end{aligned}$$

Bayes lineáris regresszió

Gyakori választás a priorra

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\mathbf{0}, \alpha^{-1}\mathbf{I})$$

Ekkor

$$\begin{aligned} \mathbf{m}_P &= \beta \mathbf{S}_P \mathbf{\Phi}^{\mathrm{T}} \mathbf{t} \\ \mathbf{S}_P^{-1} &= \alpha \mathbf{I} + \beta \mathbf{\Phi}^{\mathrm{T}} \mathbf{\Phi}. \end{aligned}$$

Bayes lineáris regresszió

Gyakori választás a priorra

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\mathbf{0}, \alpha^{-1}\mathbf{I})$$

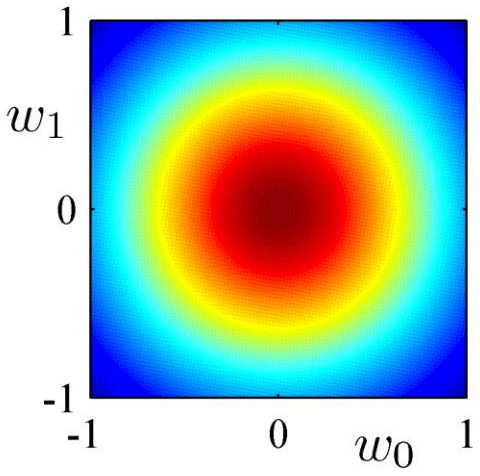
Ekkor

$$\begin{aligned} \mathbf{m}_P &= \beta \mathbf{S}_P \mathbf{\Phi}^{\mathrm{T}} \mathbf{t} \\ \mathbf{S}_P^{-1} &= \alpha \mathbf{I} + \beta \mathbf{\Phi}^{\mathrm{T}} \mathbf{\Phi}. \end{aligned}$$

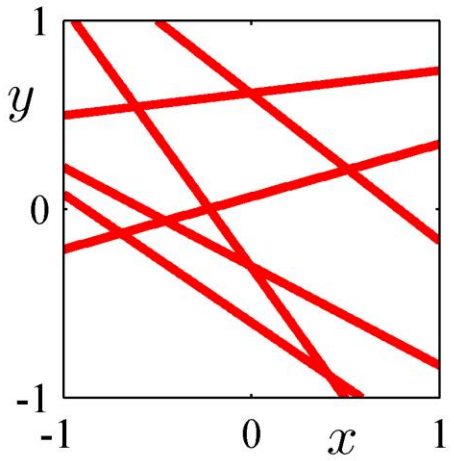
Bayes lineáris regresszió

Megfigyelések előtt $\mathbf{w}$ feltételezett eloszlása, és az ilyen $\mathbf{w}$-kkel generált adatok

Prior



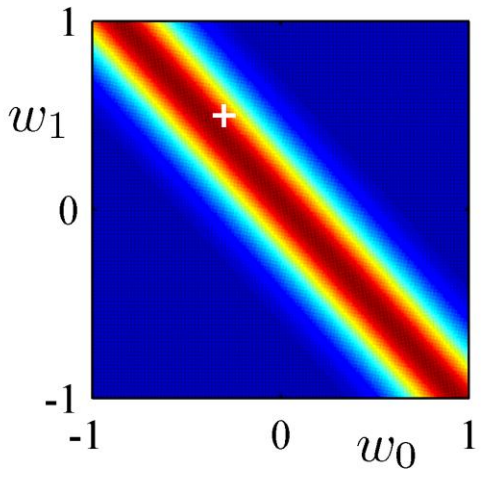
Adatok tere



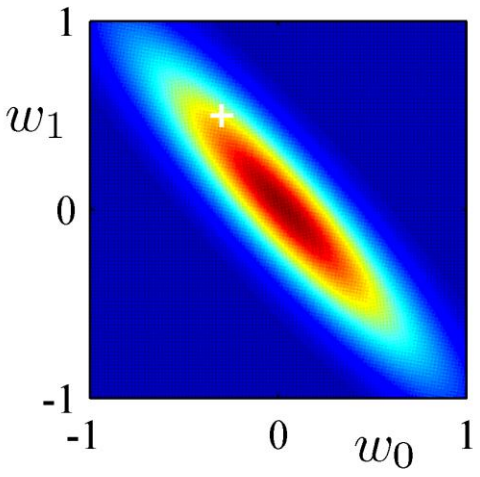
Bayes lineáris regresszió

1 megfigyelés érkezett

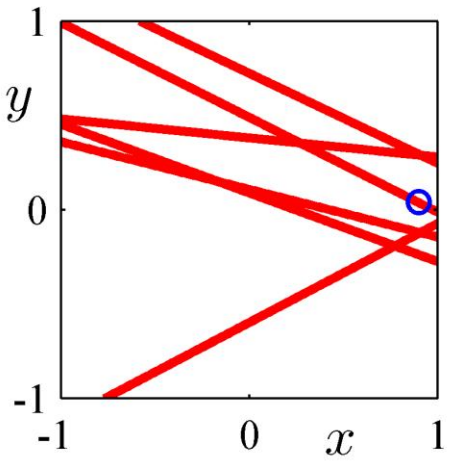
Likelihood



Poszterior



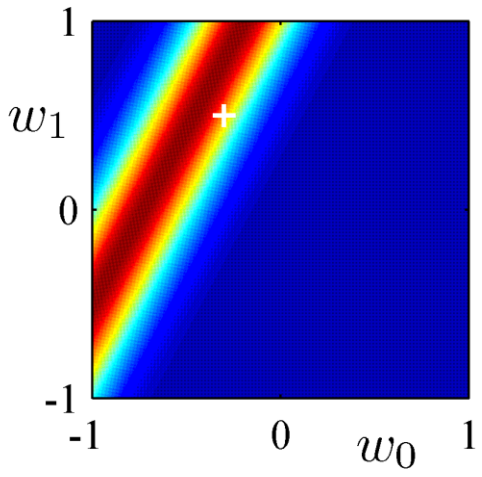
Adatok tere



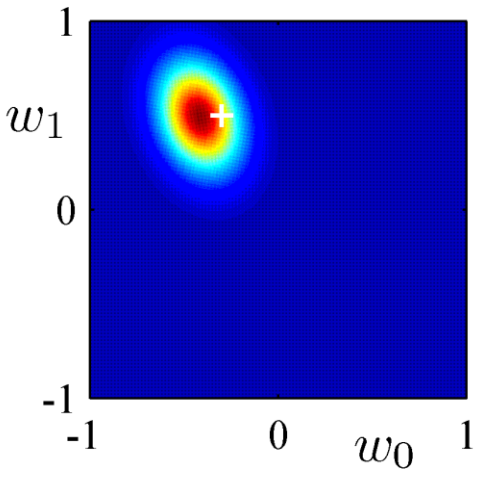
Bayes lineáris regresszió

2 megfigyelés után

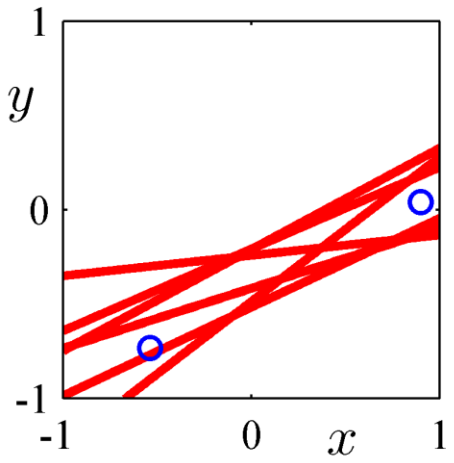
Likelihood



Poszterior



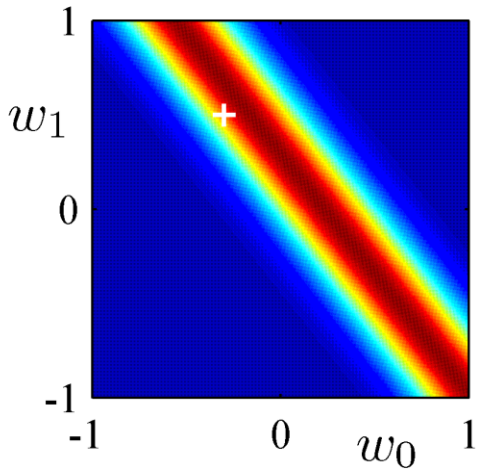
Adatok tere



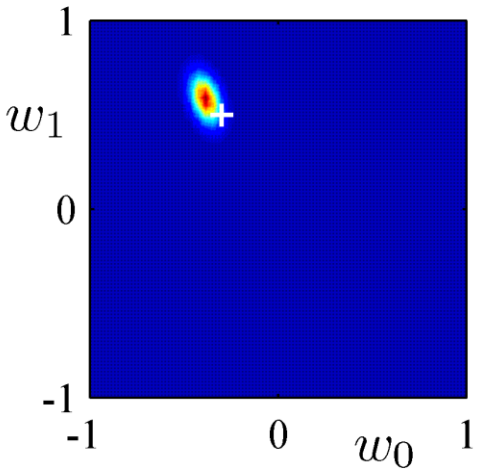
Bayes lineáris regresszió

20 megfigyelt adat után

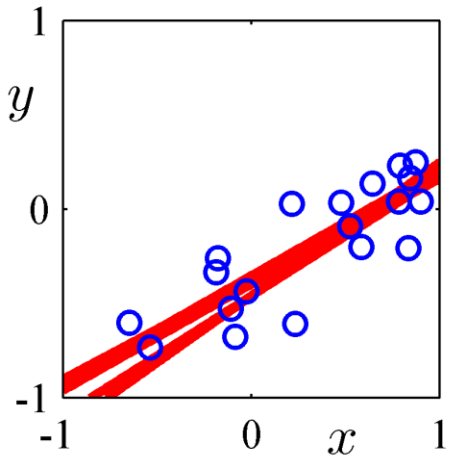
Likelihood



Poszterior



Adatok tere



Prediktív eloszlás

Jósoljuk a *d* választ egy új $\mathbf{x}$ értékre úgy hogy $\mathbf{w}$ felett integráljuk a posteriort:

$$\begin{aligned} p(d|\mathbf{d},\alpha,\beta) &= \int p(d|\mathbf{w},\beta)p(\mathbf{w}|\mathbf{d},\alpha,\beta)\,\mathrm{d}\mathbf{w} \\ &= \mathcal{N}(d|\mathbf{m}_P^{\mathrm{T}}\boldsymbol{\varphi}(\mathbf{x}),\sigma^2(\mathbf{x})) \end{aligned}$$

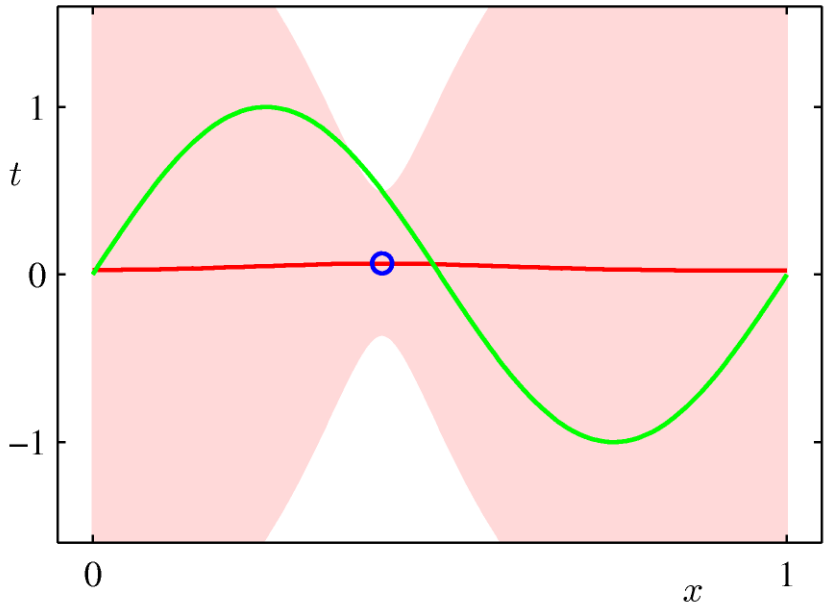
$$\sigma_P^2(\mathbf{x}) = \frac{1}{\beta} + \boldsymbol{\varphi}(\mathbf{x})^{\mathrm{T}}\mathbf{S}_P\boldsymbol{\varphi}(\mathbf{x}).$$

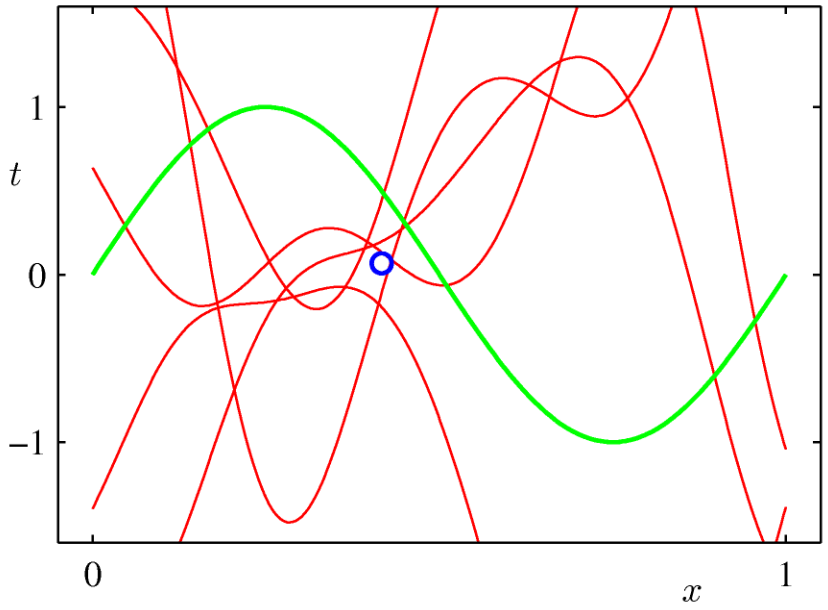
ahol

$$\begin{aligned} \mathbf{m}_P &= \beta\mathbf{S}_P\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{t} \\ \mathbf{S}_P^{-1} &= \alpha\mathbf{I} + \beta\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}. \end{aligned}$$

Prediktív eloszlás

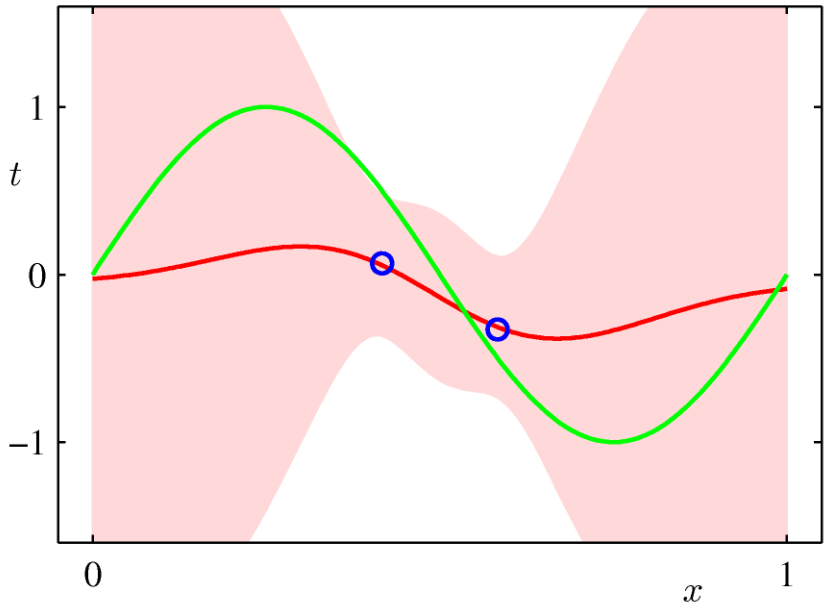
Színusz regresszió, 9 Gauss bázisfüggvény,és 1 adatpont esetén

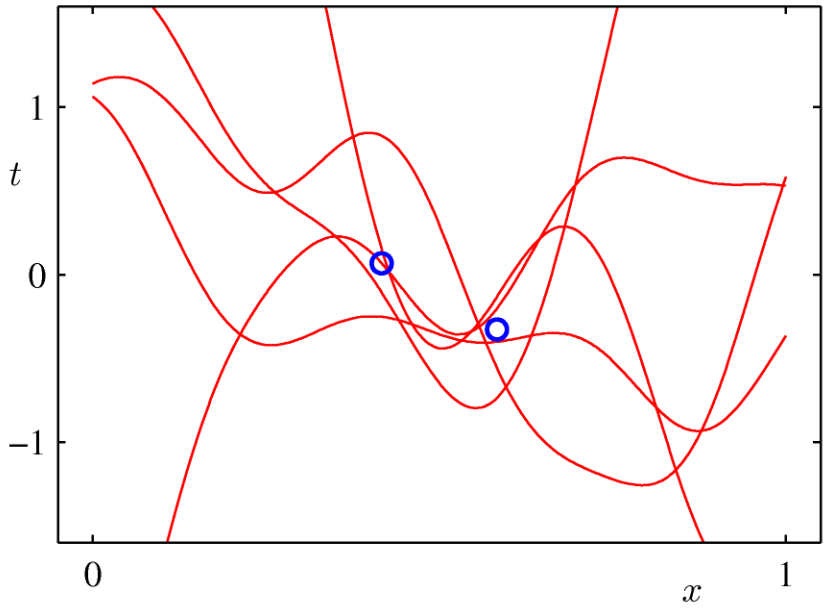




Prediktív eloszlás

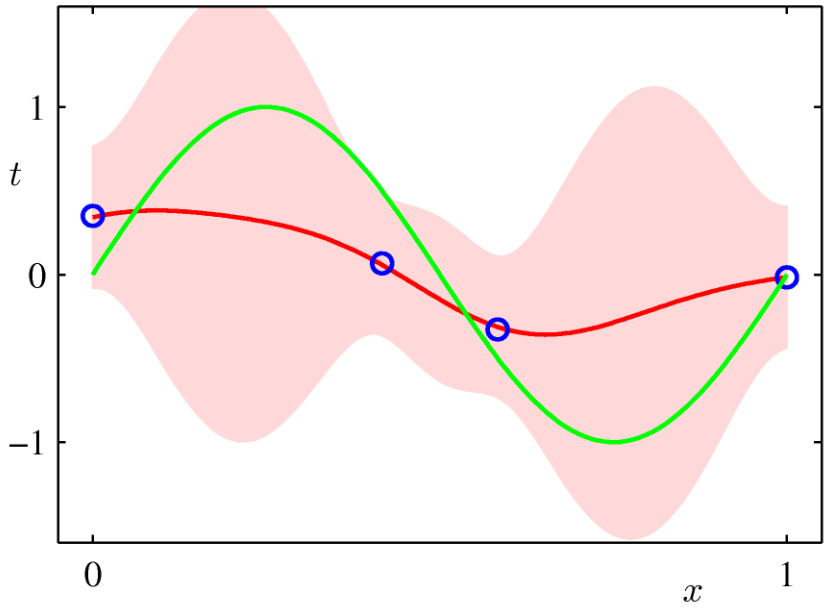
Színusz regresszió, 9 Gauss bázisfüggvény és 2 adatpont esetén

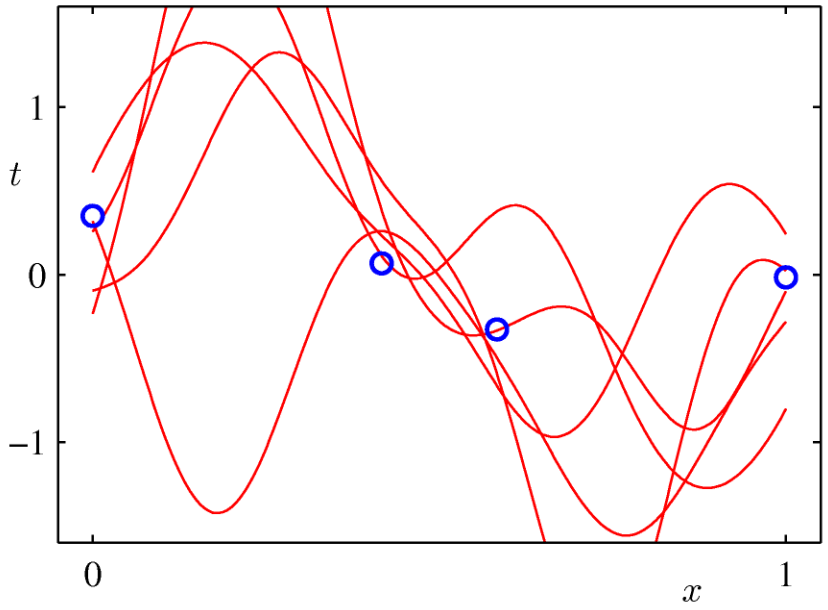




Prediktív eloszlás

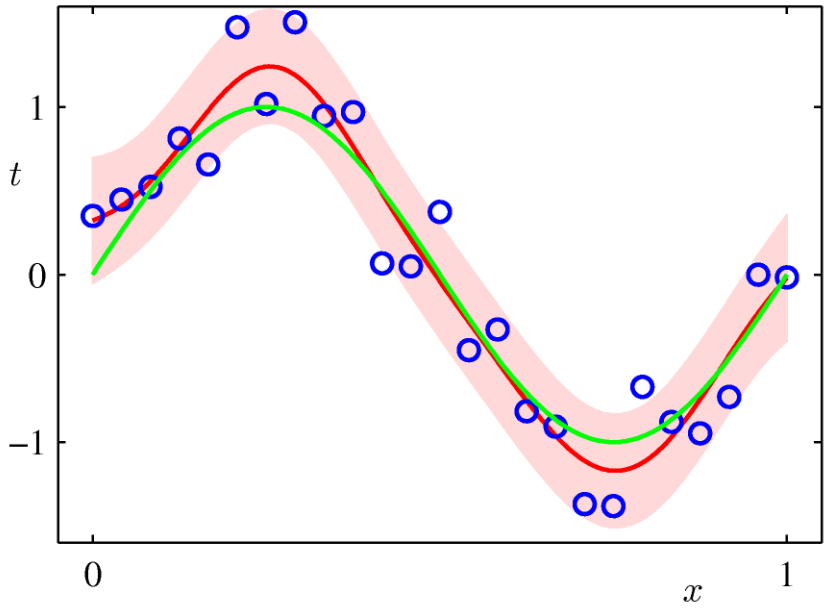
Színusz regresszió, 9 Gauss bázisfüggvény és 4 adatpont esetén

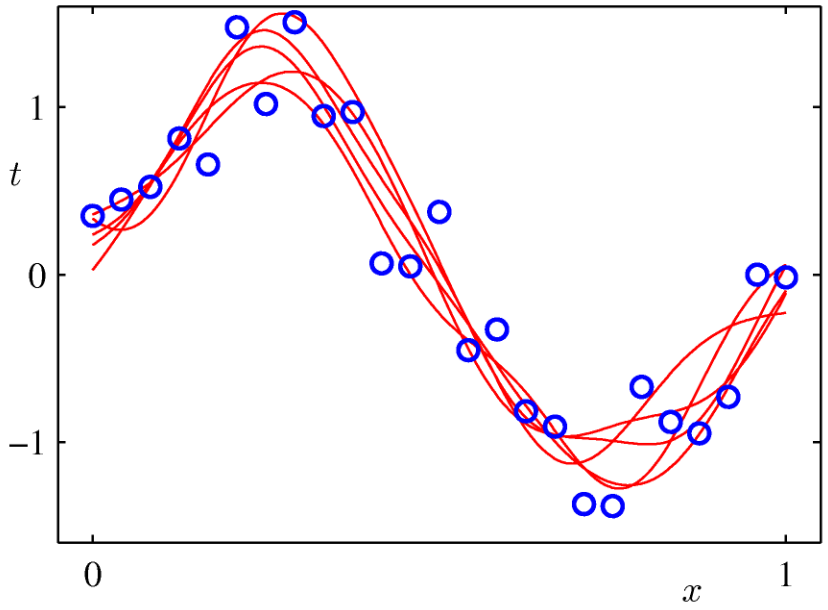




Prediktív eloszlás

Színusz regresszió, 9 Gauss bázisfüggvény és 25 adatpont esetén





Bayes modell összehasonlítás

Hogyan választhatjuk a megfelelő modellt?

Tételezzük fel, hogy van több modellünk: $\mathcal{M}_i$, $i=1, \ldots, L$, és $\mathcal{D}$ adatkészletünk; a modell posterior valószínűségét szeretnénk meghatározni

$$p(\mathcal{M}_i|\mathcal{D}) \propto p(\mathcal{M}_i)p(\mathcal{D}|\mathcal{M}_i).$$

Posterior Prior *Model evidencia vagy marginális likelihood*

Bayes faktor: a modell evidenciák hányadosa

$$\frac{p(\mathcal{D}|\mathcal{M}_i)}{p(\mathcal{D}|\mathcal{M}_j)}$$

Bayes modell összehasonlítás

Ha ismerjük $p(\mathcal{M}_i|\mathcal{D})$-ket, kiszámítható a prediktív eloszlást meghatározható, hogy az adatokból eredőben milyen jóslás adható. Most a modellek felett átlagolunk, tehát azt nézzük meg, hogy ha az adatokból többféle modell is konstruálható valamilyen poszteriorral, akkor a jósolható adat milyen eloszlású

$$p(d|\mathbf{x}, \mathcal{D}) = \sum_{i=1}^{L} p(d|\mathbf{x}, \mathcal{M}_i, \mathcal{D}) p(\mathcal{M}_i|\mathcal{D}).$$

E helyett célszerűbb azt a modellt kiválasztani, melynek az evidenciája a legnagyobb.

Bayes modell összehasonlítás

A $\mathbf{w}$ paraméterű modellnél a model evidenciát a w feletti integrálással kaphatjuk (marginalizálás)

$$p(\mathcal{D}|\mathcal{M}_i) = \int p(\mathcal{D}|\mathbf{w}, \mathcal{M}_i) p(\mathbf{w}|\mathcal{M}_i)\, \mathrm{d}\mathbf{w}.$$

Megjegyezzük, hogy

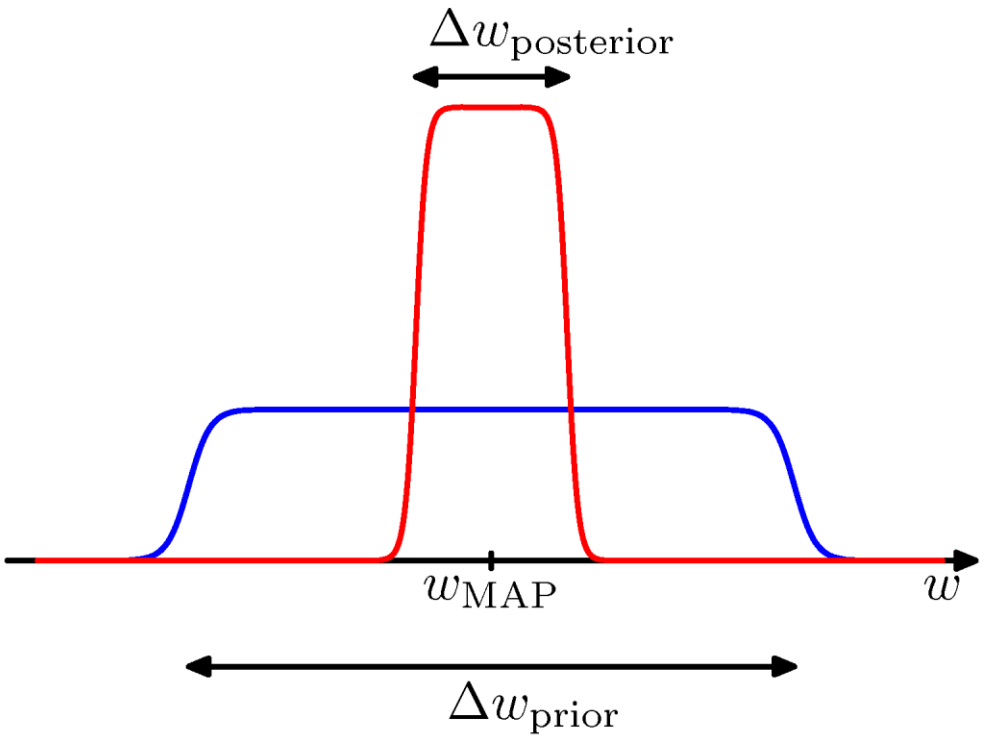
$$p(\mathbf{w}|\mathcal{D}, \mathcal{M}_i) = \frac{p(\mathcal{D}|\mathbf{w}, \mathcal{M}_i) p(\mathbf{w}|\mathcal{M}_i)}{p(\mathcal{D}|\mathcal{M}_i)}$$

Bayes modell összehasonlítás

Ha van egy adott modellünk egy skalár w paraméterrel az alábbi közelítés tehető

$$p(\mathcal{D}) = \int p(\mathcal{D}|w)p(w)\,\mathrm{d}w$$

$$\simeq \quad p(\mathcal{D}|w_{\mathrm{MAP}})\frac{\Delta w_{\mathrm{posterior}}}{\Delta w_{\mathrm{prior}}}$$



Feltételezve, hogy a posterior kis szórású (csúcsos).

Bayes model összehasonlítás

Véve a logaritmusát

$$\ln p(\mathcal{D}) \simeq \ln p(\mathcal{D}|w_{\mathrm{MAP}}) + \ln\left(\frac{\Delta w_{\mathrm{posterior}}}{\Delta w_{\mathrm{prior}}}\right).$$

Negatív

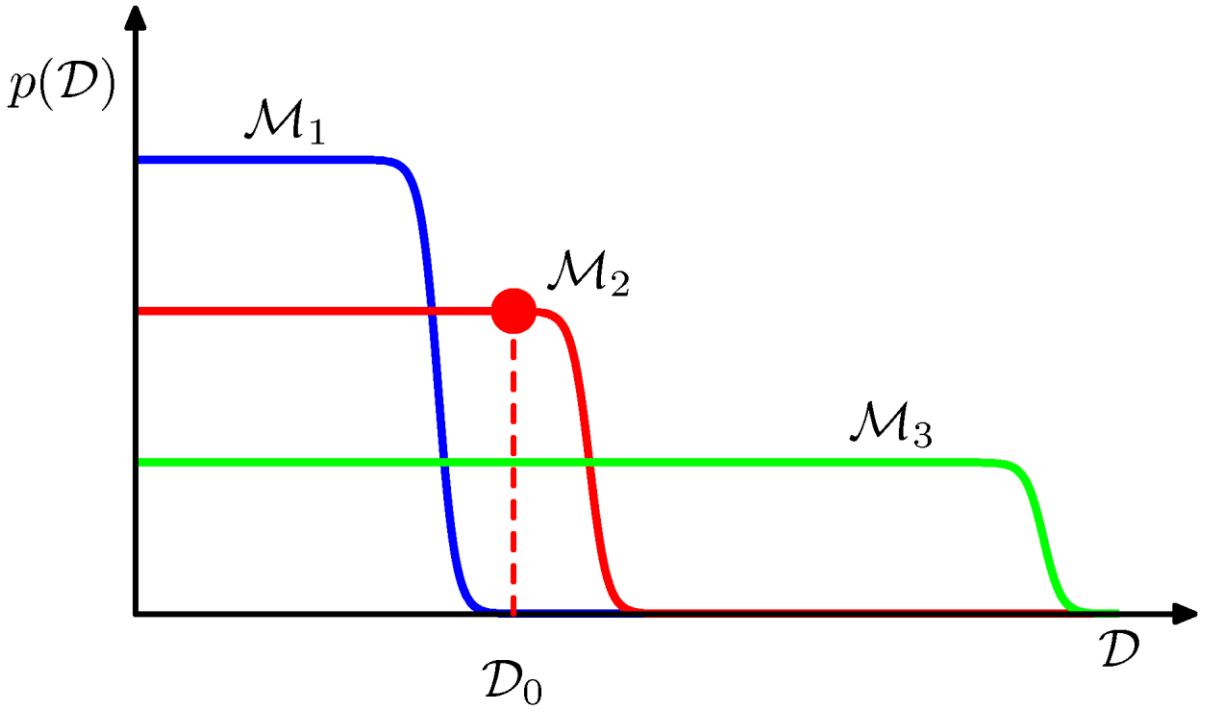
Ha nem egy, hanem M paraméterünk van, és mindegyiknél azonos $\Delta w_{\mathrm{posterior}}/\Delta w_{\mathrm{prior}}$ arányt tételezünk fel

$$\ln p(\mathcal{D}) \simeq \ln p(\mathcal{D}|\mathbf{w}_{\mathrm{MAP}}) + M \ln\left(\frac{\Delta w_{\mathrm{posterior}}}{\Delta w_{\mathrm{prior}}}\right).$$

Negatív és *M*-mel lineárisan nő a számérték.

Bayes modell összehasonlítás

Az adat és a megfelelő komplexitású modell illesztése



A prior eloszlások paramétereinek meghatározása

Meg kell határoznunk az optimális α és β paramétereket. Keressük $\ln p(\mathbf{d}|\alpha,\beta)$ maximumát α és β szerint.

$$\alpha = \frac{\gamma}{\mathbf{m}_P^{\mathrm{T}}\mathbf{m}_P}$$

$$\frac{1}{\beta} = \frac{1}{P-\gamma}\sum_{i=1}^{P}\left\{d_i - \mathbf{m}_P^{\mathrm{T}}\boldsymbol{\varphi}(\mathbf{x}_i)\right\}^2$$

Ahol

$$\gamma = \sum_i \frac{\lambda_i}{\alpha+\lambda_i}.$$

γ mind α-tól, mind β-tól függ.

A regularizációs együttható származtatása $\lambda = \frac{\alpha}{\beta}$