Bayesovská regrese

Karel Bílek

2013

1 Regrese se známou σ^2

1.1 Definice problému

Definice 1 (Model regrese se známou σ^2)**.**

- Trénovací data jsou vstupy $x_{old} = (x_1, x_2 ... x_n)^T$, výstupy $y_{old} = (y_1, y_2 ... y_n)^T$, můžou být diskrétní i spojité.
- Teď chci novému vstupu x_{new} chci přiřadit výstup y_{new}. Předpokládám model dat R, popsaný níže.
- Model dat R vypadá následovně:
  - známe σ^2
  - známe nějakou konečnou množinu bázových funkcí $\phi_i(x)$, $i = 1...M$.
  - $y_i = f(x_i) + \varepsilon_i$, kde
  - $\varepsilon_i \sim^{iid} \mathcal{N}\left(0, \sigma^2\right)$ a
  - $f(x) = \sum_{m=1}^{M} w_m \phi_m(x)$.
- Tj. model je lineární kombinace funkcí, zašuměná o nějaké ε.
- Parametry modelu jsou tedy váhy $w = (w_1, w_2 ... w_M)^T$, struktura modelu je definována množinou funkcí.
- Dále si v rámci modelu zadefinujeme priorní pravděpodobnost $w \sim \mathcal{N}(\mu_0, \Sigma_0)$
  - Tj. konstanty jsou $\mu_0, \Sigma_0, \sigma^2, \{\phi_i\}$, chápu to vše jako součást modelu R.

1.2 Pomocná lemmata

Zde si postupně napíšu a dokážu lemmata, co potom využiji ve výpočtech prediktivních funkcí dál.

Pozn. k notaci: Konstantu u výpočtu logaritmu si často budu značit c. Pokud někde napíšu $f(x) = a(x) + c = b(x) + c$, nemusí jít o stejnou konstantu, je tím prostě myšleno „něco, co nezávisí na x“.

Lemma 1. *M je symetrická matice, $a^T\, M\, b$ je skalár, potom $a^T\, M\, b = b^T\, M\, a$*

Důkaz.

- $a^T M b = \left(a^T M b\right)^T$, protože jde o skalár
- $\left(\left(a^T M\right) b\right)^T = b^T \left(a^T M\right)^T$, protože $(ab)^T = b^T a^T$
- $b^T \left(a^T M\right)^T = b^T M^T a^{T^T}$
- $b^T M^T a^{T^T} = b^T M a$, protože M je symetrická a $a^{T^T} = a$ $\square$

Lemma 2. *M je symetrická matice. Potom*

$$(a-b)^T M(a-b) = a^T Ma - 2a^T Mb + b^T Mb = a^T Ma - 2b^T Ma + b^T Mb$$

Důkaz.

- $\Big((a-b)^T\, M\Big)(a-b) = \big(a^T M - b^T M\big)(a-b) = a^T Ma - a^T Mb - b^T Ma + b^T Mb$ - „hloupé" rozepsání
- $a^T Ma - a^T Mb - b^T Ma + b^T Mb = a^T Ma - 2a^T Mb + b^T Mb$, protože M je symetrická a Lemma 1. □

Lemma 3 (Převod na čtverec)**.** *Λ je pozitivně definitní matice. Potom*

$$-\frac{1}{2}x^T \Lambda x + x^T m = -\frac{1}{2}(x - \Lambda^{-1}m)^T \Lambda (x - \Lambda^{-1}m) + \frac{1}{2}m^T \Lambda^{-1} m$$

Pozn.: tuto „zvláštní" formu používám proto, že takto se vyskytuje v exponentu $\mathcal{N}$.

Pozn. 2: Správně bych při použití tohoto lemmatu měl vždy kontrolovat, jestli jde o pozitivně definitní matici. Jelikož tohle já ale neumím, tak to dělat nebudu.
Důkaz.

- Podle Lemma 2 $(x-\Lambda^{-1}m)^T\Lambda(x-\Lambda^{-1}m) = x^T\Lambda x - 2x^T\Lambda(\Lambda^{-1}m) + (\Lambda^{-1}m)^T\Lambda(\Lambda^{-1}m)$
- protože $\Lambda\Lambda^{-1} = 1$, tak $(x-\Lambda^{-1}m)^T\Lambda(x-\Lambda^{-1}m) = x^T\Lambda x - 2x^T m + (\Lambda^{-1}m)^T m$
- Poslední člen lze přepsat jako $m^T\Lambda^{-1}m$, protože Λ^{-1} je symetrická.
- převod doleva: $(x-\Lambda^{-1}m)^T\Lambda(x-\Lambda^{-1}m) - m^T\Lambda^{-1}m = x^T\Lambda x - 2x^T m$
- Vynásobíme $-\frac{1}{2}$: $-\frac{1}{2}(x-\Lambda^{-1}m)^T\Lambda(x-\Lambda^{-1}m) + \frac{1}{2}m^T\Lambda^{-1}m = -\frac{1}{2}x^T\Lambda x + x^T m$ □

Lemma 4 (Násobení matice ze stran)**.** $\begin{pmatrix} x \\ y \end{pmatrix}^T \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} u \\ v \end{pmatrix} = x^T Au + y^T Dv + x^T Bv + y^T Cu$

Lemma 5. *Mějme*

$$f(y) = \int g(y)p(x,y)dx,$$

kde $g(y)$ je nějaká funkce nezávislá na x, a $p(x,y) \propto \bar{p}(x|y)$,

$$\bar{p}(x|y) \sim \mathcal{N}.$$

Pak $f(y) \propto g(y)$.
(Pozn.: $p(x,y) \propto \bar{p}(x|y)$ znamená, že konstanta c, přes kterou jsou proporcionální, nesmí záviset ani na x ani na y.)

Důkaz.

- $f(y) = \int g(y)p(x,y)dx = g(y)\int p(x,y)dx$ - v rámci integrálu je $g(y)$ konstanta, tak jí můžu vytknout
- $\int \bar{p}(x|y)dx = 1$
- $\int p(x,y)dx = \int c\,\bar{p}(x|y)dx = c\int \bar{p}(x|y)dx$ protože c nezávisí na x
- $c\int \bar{p}(x|y)dx = c$
- $g(y)\int p(x,y)dx = g(y)\,c \propto g(y)$ protože c nezávisí na y □

Definice 2. Nechť $x = \begin{pmatrix} x_a \\ x_b \end{pmatrix}$, $p(x) = \mathcal{N}(x|\mu,\Sigma)$. μ se potom dá napsat jako $\mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix}$. Podobně matice Σ lze rozepsat jako $\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}$

Lemma 6.
$$\begin{pmatrix} A & B \\ C & D \end{pmatrix}^{-1} = \begin{pmatrix} M & -MBD^{-1} \\ -D^{-1}CM & D^{-1} + D^{-1}CMBD^{-1} \end{pmatrix},$$
kde
$$M = \left(A - BD^{-1}C\right)^{-1}$$

Proof. Asi bych na to šel sporem. Ale udělal bych tam hromadu chyb, tak to nedělám. □

Definice 3. Inverzi kovarianční matice nazvu Λ [1], $\Lambda = \Sigma^{-1}$.

Definice 4. Λ se dá taky rozepsat jako $\Lambda = \begin{pmatrix} \Lambda_{aa} & \Lambda_{ab} \\ \Lambda_{ba} & \Lambda_{bb} \end{pmatrix}$

Lemma 7. *Platí, že* $\Lambda_{ab}^T = \Lambda_{ba}$

Důkaz.

- pozitivně definitní matice je symetrická, Λ je p.d. protože Σ je
- z toho už je to vidět □

Lemma 8 (Marginál)**.** *Definuju* $x = \begin{pmatrix} x_a \\ x_b \end{pmatrix}$, $p(x) \sim \mathcal{N}(x|\mu, \Sigma)$. *Potom platí, že marginál*

$$p(x_a) = \int p(x) dx_b \sim \mathcal{N}\left(\Sigma_{aa}, \mu_a\right)$$

Důkaz.

- $\ln p(x) = -\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu) + c$, c nezávisí na x, z definice.
- $-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu) = -\frac{1}{2}\left(x^T \Sigma^{-1} x\right) + \left(x^T \Sigma^{-1}\mu\right) - \left[\frac{1}{2}\mu^T \Sigma^{-1}\mu\right]$
  podle Lemma 2, poslední člen můžu brát jako konstantu a tedy dále ignorovat.
- $x^T \Sigma^{-1} x = x_a^T \Lambda_{aa} x_a + 2x_b^T \Lambda_{ba} x_a + x_b^T \Lambda_{bb} x_b$ (kvůli Lemmatům 1, 4, 7)
- $x^T \Sigma^{-1}\mu = x_a^T \Lambda_{aa}\mu_a + x_a^T \Lambda_{ab}\mu_b + x_b^T \Lambda_{ba}\mu_a + x_b^T \Lambda_{bb}\mu_b$
- $\ln p(x) = -\frac{1}{2}\left(x_a^T \Lambda_{aa} x_a + 2x_b^T \Lambda_{ba} x_a + x_b^T \Lambda_{bb} x_b\right) +$
  $+ \left(x_a^T \Lambda_{aa}\mu_a + x_a^T \Lambda_{ab}\mu_b + x_b^T \Lambda_{ba}\mu_a + x_b^T \Lambda_{bb}\mu_b\right) + c$ (dosazení)
- $\ln p(x) = -\frac{1}{2}x_a^T \Lambda_{aa} x_a - x_b^T \Lambda_{ba} x_a - \frac{1}{2}x_b^T \Lambda_{bb} x_b +$
  $+ x_a^T \Lambda_{aa}\mu_a + x_a^T \Lambda_{ab}\mu_b + x_b^T \Lambda_{ba}\mu_a + x_b^T \Lambda_{bb}\mu_b + c$ (roznásobení)
- $\ln p(x) = -\frac{1}{2}x_b^T \Lambda_{bb} x_b + x_b^T \left(\Lambda_{ba}\left(-x_a + \mu_a\right) + \Lambda_{bb}\mu_b\right) + \left[x_a^T \Lambda_{aa}\left(-\frac{1}{2}x_a + \mu_a\right) + x_a^T \Lambda_{ab}\mu_b\right] + c$
  (popřehazování)
- označíme si celý termín $\left(\Lambda_{ba}\left(-x_a + \mu_a\right) + \Lambda_{bb}\mu_b\right)$ jako m
  - Díváme se na část bez závorky vpravo.
  - $-\frac{1}{2}x_b^T \Lambda_{bb} x_b + x_b^T m$
  - Podle Lemmatu 3 můžu přepsat na $-\frac{1}{2}\left(x_b - \Lambda_{bb}^{-1}m\right)^T \Lambda_{bb}\left(x_b - \Lambda_{bb}^{-1}m\right) + \frac{1}{2}m^T \Lambda_{bb}^{-1} m$
- $\ln p(x_a) = -\frac{1}{2}\left(x_b - \Lambda_{bb}^{-1}m\right)^T \Lambda_{bb}\left(x_b - \Lambda_{bb}^{-1}m\right) +$
  $+ \left[\frac{1}{2}m^T \Lambda_{bb}^{-1} m + x_a^T \Lambda_{aa}\left(-\frac{1}{2}x_a + \mu_a\right) + x_a^T \Lambda_{ab}\mu_b\right]$ (viz předchozí body)
- $p\left(x_a\right) = \int c_1 \exp\left(-\frac{1}{2}\left(x_b - \Lambda_{bb}^{-1}m\right)^T \Lambda_{bb}\left(x_b - \Lambda_{bb}^{-1}m\right)\right) \cdot$
  $\exp\left(\frac{1}{2}m^T \Lambda_{bb}^{-1} m + x_a^T \Lambda_{aa}\left(-\frac{1}{2}x_a + \mu_a\right) + x_a^T \Lambda_{ab}\mu_b\right) dx_b$

[1] velká lambda. Anglicky se nazývá *precision matrix*, česky vůbec nevim

- Toto vyhovuje Lemma 5 - $p(x,y)$ z lemmatu je první a druhý člen, druhý je exponentem mult. gausiánu a liší se od něj tedy pouze normalizační konstantou, závislou na Λ_{bb}; a $g(y)$ z lemmatu je třetí člen (m nezávisí na x_b)
- tj. $p(x_a) \propto \exp\left(\frac{1}{2}m^T\Lambda_{bb}^{-1}m + x_a^T\Lambda_{aa}\left(-\frac{1}{2}x_a + \mu_a\right) + x_a^T\Lambda_{ab}\mu_b\right)$
- Jdu dělat úpravy $\frac{1}{2}m^T\Lambda_{bb}^{-1}m + x_a^T\Lambda_{aa}\left(-\frac{1}{2}x_a + \mu_a\right) + x_a^T\Lambda_{ab}\mu_b$
  - pozn: $(a-b)^T\Sigma c = c^T\Sigma(a-b) = c^T\Sigma a - c^T\Sigma b = a^T\Sigma c - b^T\Sigma c$, pokud jsou výsledky skaláry a $\Sigma^T = \Sigma$
  - $m^T\Lambda_{bb}^{-1}m = \left(\Lambda_{ba}\left(-x_a+\mu_a\right) + \Lambda_{bb}\mu_b\right)^T\Lambda_{bb}^{-1}\left(\Lambda_{ba}\left(x_a+\mu_a\right) + \Lambda_{bb}\mu_b\right) =$
    $= \left(\Lambda_{ba}\left(-x_a+\mu_a\right)\right)^T\Lambda_{bb}^{-1}\left(\Lambda_{ba}\left(-x_a+\mu_a\right)\right) +$
    $+2\left(\Lambda_{ba}\left(-x_a+\mu_a\right)\right)^T\Lambda_{bb}^{-1}\Lambda_{bb}\mu_b + \left(\Lambda_{bb}\mu_b\right)^T\Lambda_{bb}^{-1}\Lambda_{bb}\mu_b =$
    $= \left(\Lambda_{ba}\mu_a - \Lambda_{ba}x_a\right)^T\Lambda_{bb}^{-1}\left(\Lambda_{ba}\mu_a - \Lambda_{ba}x_a\right) +$
    $+2\left(\Lambda_{ba}\mu_a - \Lambda_{ba}x_a\right)^T\Lambda_{bb}^{-1}\Lambda_{bb}\mu_b + \mu_b^T\Lambda_{bb}^T\Lambda_{bb}^{-1}\Lambda_{bb}\mu_b =$
    $= \left(\Lambda_{ba}\mu_a\right)^T\Lambda_{bb}^{-1}\Lambda_{ba}\mu_a - 2\left(\Lambda_{ba}\mu_a\right)^T\Lambda_{bb}^{-1}\Lambda_{ba}x_a + \left(\Lambda_{ba}x_a\right)^T\Lambda_{bb}^{-1}\Lambda_{ba}x_a +$
    $+2\left(\Lambda_{ba}\mu_a\right)^T\Lambda_{bb}^{-1}\Lambda_{bb}\mu_b - 2\left(\Lambda_{ba}x_a\right)^T\Lambda_{bb}^{-1}\Lambda_{bb}\mu_b + \mu_b^T\Lambda_{bb}^T\Lambda_{bb}^{-1}\Lambda_{bb}\mu_b =$
    $= c_1 - 2x_a^T\left(\Lambda_{ba}^T\Lambda_{bb}^{-1}\Lambda_{ba}\mu_a\right) + x_a^T\left(\Lambda_{ba}^T\Lambda_{bb}^{-1}\Lambda_{ba}\right)x_a + c_2 - 2x_a^T\left(\Lambda_{ba}^T\Lambda_{bb}^{-1}\Lambda_{bb}\mu_b\right) + c_3 =$
    $= -2x_a^T\left(\Lambda_{ba}^T\Lambda_{bb}^{-1}\Lambda_{ba}\mu_a\right) + x_a^T\left(\Lambda_{ba}^T\Lambda_{bb}^{-1}\Lambda_{ba}\right)x_a - 2x_a^T\left(\Lambda_{ba}^T\mu_b\right) + c_4$
    $= -2x_a^T\left(\Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba}\mu_a\right) + x_a^T\left(\Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba}\right)x_a - 2x_a^T\left(\Lambda_{ab}\mu_b\right) + c_4$
    kde c_n jsou konstanty nezávislé na x_a
  - $\frac{1}{2}m^T\Lambda_{bb}^{-1}m + x_a^T\Lambda_{aa}\left(-\frac{1}{2}x_a + \mu_a\right) + x_a^T\Lambda_{ab}\mu_b =$
    $= \frac{1}{2}\left(-2x_a^T\left(\Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba}\mu_a\right) + x_a^T\left(\Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba}\right)x_a - 2x_a^T\left(\Lambda_{ab}\mu_b\right)\right) +$
    $+x_a^T\Lambda_{aa}\left(-\frac{1}{2}x_a + \mu_a\right) + x_a^T\Lambda_{ab}\mu_b + c_5 =$
    $= -x_a^T\left(\Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba}\mu_a\right) + x_a^T\left(\frac{1}{2}\Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba}\right)x_a - x_a^T\left(\Lambda_{ab}\mu_b\right) +$
    $+x_a^T\left(-\frac{1}{2}\Lambda_{aa}\right)x_a + x_a^T\left(\Lambda_{aa}\mu_a\right) + x_a^T\left(\Lambda_{ab}\mu_b\right) + c_5 =$
    $= -x_a^T\left(\Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba}\mu_a\right) + x_a^T\left(\frac{1}{2}\Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba}\right)x_a +$
    $+x_a^T\left(-\frac{1}{2}\Lambda_{aa}\right)x_a + x_a^T\left(\Lambda_{aa}\mu_a\right) + c_5 =$
    $= x_a^T\left(\Lambda_{aa}\mu_a - \Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba}\mu_a\right) + x_a^T\left(\frac{1}{2}\Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba} - \frac{1}{2}\Lambda_{aa}\right)x_a + c_5 =$
    $= -\frac{1}{2}x_a^T\left(\Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba}\right)x_a + x_a^T\left(\left(\Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba}\right)\mu_a\right) + c_5$
- podle Lemma 3 a definice $\mathcal{N}$ tedy platí, že $p\left(x_a\right) \sim \mathcal{N}\left(\left(\Lambda_{aa} - \Lambda_{ab}\Lambda_{bb}^{-1}\Lambda_{ba}\right)^{-1}, \mu_a\right)$
- podle Lemma 6 je to ve skutečnosti $p\left(x_a\right) \sim \mathcal{N}\left(\Sigma_{aa}, \mu_a\right)$ □

Lemma 9 (Marginál násobku)**.** *Pokud platí* $p(x) = \mathcal{N}\left(x|\mu, \Lambda^{-1}\right)$, $p(y|x) = \mathcal{N}\left(y|Ax+b, L^{-1}\right)$, *tak poté*

$$p(y) = \int p(y|x)p(x)dx \sim \mathcal{N}\left(A\mu + b, L^{-1} + A\Lambda^{-1}A^T\right)$$

Důkaz.

- Podíváme-li se na Lemma 8 a definujeme si například $z = \begin{pmatrix} x \\ y \end{pmatrix}$, vidíme, že jde o stejný případ.

  Tj. musíme spočítat matici a rozptyl $p(y|x)p(x)$ a „výsek“, odpovídající y, je potřebný výsledek.
- $\ln p(z) = \ln p(x) + \ln p(y|x)$ - definice
- $\ln p(z) = -\frac{1}{2}(x-\mu)^T\Lambda(x-\mu) - \frac{1}{2}(y - Ax - b)^T L(y - Ax - b) + c$
  - $(x-\mu)^T\Lambda(x-\mu) = x^T\Lambda x - 2x^T\Lambda\mu + c$
  - $(y - (Ax+b))^T L(y - (Ax+b)) = y^T Ly - 2(Ax+b)^T Ly + (Ax+b)^T L(Ax+b) =$
    $= y^T Ly - 2x^T A^T Ly - 2y^T Lb + x^T A^T LAx + 2x^T LAb + c$

– $\ln p(z) = -\frac{1}{2}x^T\Lambda x + x^T\Lambda\mu - \frac{1}{2}y^TLy + x^TA^TLy + y^TLb - \frac{1}{2}x^TA^TLAx - x^TLAb + c =$
$= \left[-\frac{1}{2}x^T\left(\Lambda + A^TLA\right)x - \frac{1}{2}y^TLy + \frac{1}{2}y^TLAx + \frac{1}{2}x^TA^TLy\right] + \left[x^T\left(\Lambda\mu - A^TLb\right) + y^TLb\right]$

– první závorka:

- $-\frac{1}{2}x^T\left(\Lambda + A^TLA\right)x - \frac{1}{2}y^TLy + \frac{1}{2}y^TLAx + \frac{1}{2}x^TA^TLy =$
  $= -\frac{1}{2}\begin{pmatrix}x\\y\end{pmatrix}^T\begin{pmatrix}\Lambda + A^TLA & -A^TL\\ -LA & L\end{pmatrix}\begin{pmatrix}x\\y\end{pmatrix}$
- chci $\begin{pmatrix}\Lambda + A^TLA & -A^TL\\ -LA & L\end{pmatrix}^{-1}$, chci použít Lemma 6
- $M = \left(\Lambda + A^TLA - A^TL\,(L)^{-1}\,LA\right)^{-1} = \Lambda^{-1}$
- $\begin{pmatrix}\Lambda + A^TLA & -A^TL\\ -LA & L\end{pmatrix}^{-1} =$
  $= \begin{pmatrix}\Lambda^{-1} & -\Lambda^{-1}\left(-A^TL\right)L^{-1}\\ -L^{-1}\left(-LA\right)\Lambda^{-1} & L^{-1} + L^{-1}\left(-LA\right)\Lambda^{-1}\left(-A^TL\right)L^{-1}\end{pmatrix} =$
  $= \begin{pmatrix}\Lambda^{-1} & \Lambda^{-1}A^T\\ A\Lambda^{-1} & L^{-1} + A\Lambda^{-1}A^T\end{pmatrix}$
- první závorka je tedy $-\frac{1}{2}\begin{pmatrix}x\\y\end{pmatrix}^T\begin{pmatrix}\Lambda^{-1} & \Lambda^{-1}A^T\\ A\Lambda^{-1} & L^{-1} + A\Lambda^{-1}A^T\end{pmatrix}^{-1}\begin{pmatrix}x\\y\end{pmatrix}$

– druhá závorka jde napsat jako $\begin{pmatrix}x\\y\end{pmatrix}^T\begin{pmatrix}\Lambda\mu - A^TLb\\ Lb\end{pmatrix}$

– tj. $ln(p) = -\frac{1}{2}\begin{pmatrix}x\\y\end{pmatrix}^T\begin{pmatrix}\Lambda^{-1} & \Lambda^{-1}A^T\\ A\Lambda^{-1} & L^{-1} + A\Lambda^{-1}A^T\end{pmatrix}^{-1}\begin{pmatrix}x\\y\end{pmatrix} + \begin{pmatrix}x\\y\end{pmatrix}^T\begin{pmatrix}\Lambda\mu - A^TLb\\ Lb\end{pmatrix} + c$

– Použijeme Lemma 3 pro převod na exponent gausiánu

– A nakonec použijeme Lemma 8.

□

Lemma 10. *Mějme $x_i \sim \mathcal{N}(\bar{x}_i, \sigma^2)\forall i \in J$ pro nějaké konečné J ($\mathcal{N}$ je tady univariate gaussian, x_i skalár). Pak $\left(x_1, x_2, ..., x_{|J|}\right)^T \sim \mathcal{N}\left(\left(\bar{x}_1, \bar{x}_2, ..., \bar{x}_{|J|}\right)^T, \sigma^2 I\right)$ (kde I je jednotková matice a $\mathcal{N}$ je multivariate gaussián)*

Důkaz.

– x jsou na sobě nezávislé.

– Tedy $p\left(\left(x_1, x_2, ..., x_{|J|}\right)^T\right) = p(x_1)p(x_2)p(x_3)...$

– $p(x_j) \propto \exp\left(-\frac{1}{2}\left(x_j - \bar{x}^2_{|J|}\sigma^2\right)\right)$ z definice

– tedy $p\left(\left(x_1, x_2, ..., x_{|J|}\right)^T\right) \propto \sum_j -\frac{1}{2}(x_j - \bar{x}_j)^2\sigma^2$

– $\sum_j -\frac{1}{2}(x_j - \bar{x}_j)^2\sigma^2 = -\frac{1}{2}\sigma^2\sum_j(x_j - \bar{x}_j)^2 =$
$= -\frac{1}{2}\sigma^2\left(\left(x_1, x_2, ..., x_{|J|}\right) - \left(\bar{x}_1, \bar{x}_2, ..., \bar{x}_{|J|}\right)\right)\left(\left(x_1, x_2, ..., x_{|J|}\right) - \left(\bar{x}_1, \bar{x}_2, ..., \bar{x}_{|J|}\right)\right)^T$
$= -\frac{1}{2}\sigma^2\left(\left(x_1, x_2, ..., x_{|J|}\right)^T - \left(\bar{x}_1, \bar{x}_2, ..., \bar{x}_{|J|}\right)^T\right)^T\left(\left(x_1, x_2, ..., x_{|J|}\right)^T - \left(\bar{x}_1, \bar{x}_2, ..., \bar{x}_{|J|}\right)^T\right)$
$= -\frac{1}{2}\sigma^2\left(\left(x_1, x_2, ..., x_{|J|}\right)^T - \left(\bar{x}_1, \bar{x}_2, ..., \bar{x}_{|J|}\right)^T\right)^T I\left(\left(x_1, x_2, ..., x_{|J|}\right)^T - \left(\bar{x}_1, \bar{x}_2, ..., \bar{x}_{|J|}\right)^T\right)$
$= -\frac{1}{2}\left(\left(x_1, x_2, ..., x_{|J|}\right)^T - \left(\bar{x}_1, \bar{x}_2, ..., \bar{x}_{|J|}\right)^T\right)^T\sigma^2 I\left(\left(x_1, x_2, ..., x_{|J|}\right)^T - \left(\bar{x}_1, \bar{x}_2, ..., \bar{x}_{|J|}\right)^T\right)$
a z toho už nám věta plyne

□

1.3 Zpátky k problému - prediktivní distribuce

- Hlavní, co chceme zjistit, je prediktivní distrubuce. Tj.

$$p(y_{new}|y_{old}, x_{old}, x_{new}, R),$$

kde R je model, co byl popsán na začátku (v R jsou „zahrnuty“ konstanty modelu). Pro jednodušší zápis nebudu podmínění R do dalších vzorečků psát.

Definice 5. Maticí Φ_{old} myslím matici $n \cdot M$, kde n je počet funkcí ϕ, M je velikost vektoru x_{old}, $\Phi_{old\,i,j} = \phi_i(x_{old\,j})$. Totéž s Φ_{new}.

Lemma 11 (Přepis na marginál)**.**

$$p(y_{new}|y_{old}, x_{old}, x_{new}) = \int p(y_{new}|w, x_{new}) p(w|y_{old}, x_{old}) dw$$

Důkaz.

- $p(y_{new}|y_{old}, x_{old}, x_{new}) = \int p(y_{new}|w, y_{old}, x_{old}, x_{new}) p(w|y_{old}, x_{old}, x_{new}) dw$
- $p(y_{new}|w, y_{old}, x_{old}, x_{new}) = p(y_{new}|w, x_{new})$ - pokud už mám váhy, tak výsledek nezávisí na ničem dalším
- $p(w|y_{old}, x_{old}, x_{new}) = p(w|y_{old}, x_{old})$, protože nová data bez zdroje nám o váhách nic neřeknou □

Lemma 12 (Přepis posterioru)**.**

$$p(w|y_{old}, x_{old}) \propto p(w) p(y_{old}|x_{old}, w)$$

Pozn.: $p(w|y_{old}, x_{old})$ je bayesovský posterior, $p(w)$ je prior, $p(y_{old}|x_{old}, w)$ je likelihood

Důkaz.

- Z Bayesova pravidla $p(w|y_{old}, x_{old}) = \frac{p(w|x_{old}) p(y_{old}|w, x_{old})}{p(y_{old}|x_{old})}$
- Z hlediska w je $p(y_{old}|x_{old})$ konstanta
- $p(w|x_{old}) = p(w)$ - bez informace o y_{old} nám x_{old} nic neřekne □

Lemma 13 (Likelihood)**.**

$$y_{new}|w, x_{new} \sim \mathcal{N}(\Phi_{new} w, \sigma^2 I),$$

kde I je jednotková matice. Podobně $y_{old}|w, x_{old} \sim \mathcal{N}(\Phi_{old} w, \sigma^2 I)$

Důkaz.

- pro y_{new} a y_{old} je následující postup stejný - prostě máme vstup a váhy a chceme pravděpodobnost výstupu. Tj. budu psát jen x a y (resp x_i a y_i)
- $f(x)$ si nazvu vektor $\left(f(x_1), f(x_2), ..., f\left(x_{|x|}\right)\right)^T$
- podle modelu $y_i = f(x_i) + \varepsilon_i \forall i \in \{1...|y|\}$, tj. $y_i \sim \mathcal{N}\left(f(x_i), \sigma^2\right)$
- To jde napsat také jako $y \sim \mathcal{N}(f(x), \sigma^2 I)$ podle Lemma 10
- $f(x)_i = f(x_i) = \sum_{m=1}^{M} w_m \phi_m(x_i)$, - to je totéž, co $[\Phi w]_i$ □

Lemma 14 (Výpočet posterioru)**.** $w|y_{old}, x_{old} \sim \mathcal{N}(\hat{w}, \hat{\Sigma})$, *kde*

$$\hat{\Sigma} = \left(\Sigma_0^{-1} + \sigma^{-2} \Phi_{old}^T \Phi_{old}\right)^{-1},$$

$$\hat{w} = \hat{\Sigma}\left(\Sigma_0^{-1} \mu_0 + \sigma^{-2} \Phi_{old}^T y_{old}\right)$$

Důkaz.

- Podle Lemmatu 12 víme, že $p(w|y_{old}, x_{old}) \propto p(w)p(y_{old}|x_{old}, w)$
- Podle Definice 1 $w \sim \mathcal{N}(\mu_0, \Sigma_0)$, tj $\ln p(w) = \left(-\frac{1}{2}(w-\mu_0)^T \Sigma_0^{-1}(w-\mu_0)\right) + c$, to lze přepsat podle Lemma 2 jako $-\frac{1}{2}w^T\Sigma_0^{-1}w + w^T\Sigma_0^{-1}\mu_0 + \left[-\frac{1}{2}\ \mu_0^T\Sigma_0^{-1}\mu_0 + c\right]$, poslední závorka je konstanta
- Podle Lemma 13 $y_{old}|x_{old}, w \sim \mathcal{N}(\Phi_{old}w, \sigma^2 I)$, tj.
  $\ln p(y_{old}|x_{old}, w) = \left(-\frac{1}{2}(y_{old}-\Phi_{old}w)^T\left(\sigma^2 I\right)^{-1}(y_{old}-\Phi_{old}w)\right) + c =$
  $= \left(-\frac{1}{2}\sigma^{-2}(y_{old}-\Phi_{old}w)^T(y_{old}-\Phi_{old}w)\right) + c =$
  $= \left(-\frac{1}{2}\sigma^{-2}\left(y_{old}^T y_{old} - 2w^T\Phi_{old}^T y_{old} + w^T\Phi_{old}^T\Phi_{old}w\right)\right) + c$
- tedy $\ln p(w|y_{old}, x_{old}) =$
  $\left(-\frac{1}{2}w^T\Sigma_0^{-1}w + w^T\Sigma_0^{-1}\mu_0\right) + \left(-\frac{1}{2}\sigma^{-2}\left(y_{old}^T y_{old} - 2w^T\Phi_{old}^T y_{old} + w^T\Phi_{old}^T\Phi_{old}w\right)\right) + c$
- V tomto okamžiku už můžeme brát y_{old} jako konstantu, neboť nezávisí na w.
- $\ln p(w|y_{old}, x_{old}) =$
  $\left(-\frac{1}{2}w^T\Sigma_0^{-1}w + w^T\Sigma_0^{-1}\mu_0\right) + \left(-\frac{1}{2}\sigma^{-2}\left(-2w^T\Phi_{old}^T y_{old} + w^T\Phi_{old}^T\Phi_{old}w\right)\right) + c$
  $= -\frac{1}{2}w^T\Sigma_0^{-1}w + w^T\Sigma_0^{-1}\mu_0 + w^T\sigma^{-2}\Phi_{old}^T y_{old} - \frac{1}{2}\sigma^{-2}w^T\Phi_{old}^T\Phi_{old}w + c$
- Opět chceme převést do formy Lemma 3.
- $\ln p(w|y_{old}, x_{old}) =$
  $= -\frac{1}{2}w^T\left(\Sigma_0^{-1} + \sigma^{-2}\Phi_{old}^T\Phi_{old}\right)w + w^T\left(\Sigma_0^{-1}\mu_0 + \sigma^{-2}\Phi_{old}^T y_{old}\right) + c =$
  $= -\frac{1}{2}w^T\hat{\Sigma}^{-1}w + w^T\left(\hat{\Sigma}^{-1}\hat{w}\right)$
- Použijeme Lemma 3. □

Věta 1 (Predictive)**.**

$$y_{new}|y_{old}, x_{old}, x_{new} \sim \mathcal{N}\left(\Phi_{new}\hat{w}, \sigma^2 I + \Phi_{new}\hat{\Sigma}\Phi_{new}^T\right)$$

Důkaz.

- Podle Lemmatu 11 víme, že $p(y_{new}|y_{old}, x_{old}, x_{new}) = \int p(y_{new}|w, x_{new})p(w|y_{old}, x_{old})dw$
- Podle Lemmatu 13 víme, že $y_{new}|w, x_{new} \sim \mathcal{N}(\Phi_{new}w, \sigma^2 I)$
- Podle Lemmatu 14 víme, že $w|y_{old}, x_{old} \sim \mathcal{N}(\hat{w}, \hat{\Sigma})$
- Použijeme Lemma 9, kde přepíšeme
  - $y \to y_{new}$
  - $x \to w$
  - $\mu \to \hat{w}$
  - $\Lambda \to \hat{\Sigma}^{-1}$
  - $L \to \sigma^{-2}I$
  - $b \to 0$
  - $A \to \Phi_{new}$
  - $\mathcal{N}\left(A\mu + b, L^{-1} + A\Lambda^{-1}A^T\right) \to \mathcal{N}\left(\Phi_{new}\hat{w}, \sigma^2 I + \Phi_{new}\hat{\Sigma}\Phi_{new}^T\right)$

□

2 Regrese s neznámou σ^2

2.1 Pomocné definice a lemmata

Definice 6 (Gamma funkce)**.** $\Gamma(z) = \int_0^\infty t^{z-1} e^{-t}\, dt$

Definice 7 (Gamma distribuce)**.**

$$\mathrm{Gam}(x|a,b) = \frac{1}{\Gamma(a)} b^a x^{a-1} \exp(-bx)$$

Lemma 15. $\mathrm{Gam}(x|a,b) \propto x^{a-1} \exp(-bx)$

Definice 8 (Normal-Gamma)**.**

$$\mathcal{NG}(w,\lambda|\mu_0,\Sigma_0,a,b) = \mathcal{N}(w|\mu_0,\lambda^{-1}\Sigma_0) \cdot \mathrm{Gam}(\lambda|a,b)$$

kde w je vektor, λ je skalár, μ_0 je vektor, Σ_0 je matice, a a b jsou skaláry.
Pozn: Tato distribuce je na různých místech definována různě, já použiji tuto definici.

Lemma 16.

$$\ln\left(\mathcal{NG}\left(w,\lambda|\mu_0,\Sigma_0,a,b\right)\right) = \ln\left(\lambda^{a+\frac{|w|-2}{2}}\right)\left(-\frac{1}{2}\lambda\left(\left(w-\mu_0\right)^T \Sigma_0^{-1}\left(w-\mu_0\right)+2b\right)\right)+c$$

Důkaz.

- $\mathcal{N}(w|\mu_0,\lambda^{-1}\Sigma_0) = (2\pi)^{-\frac{|w|}{2}} \left|\lambda^{-1}\Sigma_0\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\left(w-\mu_0\right)^T \Sigma_0^{-1}\lambda\left(w-\mu_0\right)\right)\right) =$
  $= (2\pi)^{-\frac{|w|}{2}} \left(\lambda^{-|w|}\left|\Sigma_0\right|\right)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\left(\left(w-\mu_0\right)^T \Sigma_0^{-1}\lambda\left(w-\mu_0\right)\right)\right) =$
  $= (2\pi)^{-\frac{|w|}{2}} \left|\Sigma_0\right|^{-\frac{1}{2}} \lambda^{\frac{|w|}{2}} \exp\left(-\frac{1}{2}\left(\left(w-\mu_0\right)^T \Sigma_0^{-1}\lambda\left(w-\mu_0\right)\right)\right) =$
  $= C\lambda^{\frac{|w|}{2}} \exp\left(-\frac{1}{2}\lambda\left(\left(w-\mu_0\right)^T \Sigma_0^{-1}\left(w-\mu_0\right)\right)\right)$
- (pozor, λ teď není konstanta)
- $\mathrm{Gam}(\lambda|a,b) = C\lambda^{a-1}\exp(-b\lambda)$
- $\mathcal{NG}\left(w,\lambda|\mu_0,\Sigma_0,a,b\right) = C\lambda^{a+\frac{|w|-2}{2}} \exp\left(-\frac{1}{2}\lambda\left(\left(w-\mu_0\right)^T \Sigma_0^{-1}\left(w-\mu_0\right)+2b\right)\right)$ □

Lemma 17. *Když Σ je poz. def. matice, tak*

$$(y-X\beta)^T(y-X\beta)+(\beta-\beta_0)\Sigma^{-1}(\beta-\beta_0)+2b = (\beta-\beta^*)^T(\Sigma^*)^{-1}(\beta-\beta^*)+2b^*$$

kde

$$\beta^* = (\Sigma^{-1}+X^TX)^{-1}(\Sigma^{-1}\beta_0+X^Ty)$$
$$\Sigma^* = (\Sigma^{-1}+X^TX)^{-1}$$
$$b^* = b+\frac{1}{2}(\beta_0^T\Sigma^{-1}\beta_0+y^Ty-(\beta^*)^T(\Sigma^*)^{-1}\beta^*)$$

Důkaz.

- Vezmu nejdřív levou stranu a rozpočítám jí, pak pravou a rozpočítám jí a obě vyjdou stejně.
- $(y-X\beta)^T(y-X\beta)+(\beta-\beta_0)\Sigma^{-1}(\beta-\beta_0)+2b =$
  $= y^Ty-2y^TX\beta+\beta^TX^TX\beta+\beta^T\Sigma^{-1}\beta-2\beta_0^T\Sigma^{-1}\beta+\beta_0^T\Sigma^{-1}\beta_0+2b$
- (Jak jsem psal někde na začátku - protože neumím dokazovat pozitivní definitnost, tak jí nedokazuju a předpokládám jí u Σ^*)

- $(\beta-\beta^*)^T(\Sigma^*)^{-1}(\beta-\beta^*) =$
  $= (\beta-(\Sigma^{-1}+X^TX)^{-1}(\Sigma^{-1}\beta_0+X^Ty))^T((\Sigma^{-1}+X^TX)^{-1})^{-1}\cdot$
  $\cdot(\beta-(\Sigma^{-1}+X^TX)^{-1}(\Sigma^{-1}\beta_0+X^Ty)) =$
  $= (\beta-(\Sigma^{-1}+X^TX)^{-1}(\Sigma^{-1}\beta_0+X^Ty))^T(\Sigma^{-1}+X^TX)\cdot$
  $\cdot(\beta-(\Sigma^{-1}+X^TX)^{-1}(\Sigma^{-1}\beta_0+X^Ty)) =$
  $= \beta^T(\Sigma^{-1}+X^TX)\beta - 2\beta^T(\Sigma^{-1}+X^TX)(\Sigma^{-1}+X^TX)^{-1}(\Sigma^{-1}\beta_0+X^Ty)+$
  $+((\Sigma^{-1}+X^TX)^{-1}(\Sigma^{-1}\beta_0+X^Ty))^T(\Sigma^{-1}+X^TX)(\Sigma^{-1}+X^TX)^{-1}(\Sigma^{-1}\beta_0+X^Ty) =$
  $= \beta^T(\Sigma^{-1}+X^TX)\beta - 2\beta^T(\Sigma^{-1}\beta_0+X^Ty)+$
  $+((\Sigma^{-1}+X^TX)^{-1}(\Sigma^{-1}\beta_0+X^Ty))^T(\Sigma^{-1}\beta_0+X^Ty) =$
  $= \beta^T(\Sigma^{-1}+X^TX)\beta - 2\beta^T(\Sigma^{-1}\beta_0+X^Ty)+$
  $+(\Sigma^{-1}\beta_0+X^Ty)^T((\Sigma^{-1}+X^TX)^{-1})^T(\Sigma^{-1}\beta_0+X^Ty) =$
  $= \beta^T\Sigma^{-1}\beta+\beta^TX^TX\beta-2\beta^T\Sigma^{-1}\beta_0-2\beta^TX^Ty+$
  $+(\Sigma^{-1}\beta_0+X^Ty)^T((\Sigma^{-1}+X^TX)^{-1})^T(\Sigma^{-1}\beta_0+X^Ty) =$
  $= \beta^T\Sigma^{-1}\beta+\beta^TX^TX\beta-2\beta^T\Sigma^{-1}\beta_0-2\beta^TX^Ty+$
  $+\beta_0^T{\Sigma^{-1}}^T((\Sigma^{-1}+X^TX)^{-1})^T\Sigma^{-1}\beta_0+2y^TX((\Sigma^{-1}+X^TX)^{-1})^T\Sigma^{-1}\beta_0+$
  $+y^TX((\Sigma^{-1}+X^TX)^{-1})^TX^Ty =$
  $= \beta^T\Sigma^{-1}\beta+\beta^TX^TX\beta-2\beta^T\Sigma^{-1}\beta_0-2\beta^TX^Ty+$
  $+\beta_0^T\Sigma^{-1}(\Sigma^{-1}+X^TX)^{-1}\Sigma^{-1}\beta_0+2y^TX(\Sigma^{-1}+X^TX)^{-1}\Sigma^{-1}\beta_0+$
  $+y^TX(\Sigma^{-1}+X^TX)^{-1}X^Ty =$
  $= \beta^T\Sigma^{-1}\beta+\beta^TX^TX\beta-2\beta^T\Sigma^{-1}\beta_0-2\beta^TX^Ty+$
  $+\beta_0^T\Sigma^{-1}\Sigma^*\Sigma^{-1}\beta_0+2y^TX\Sigma^*\Sigma^{-1}\beta_0+$
  $+y^TX\Sigma^*X^Ty$
- $2b^* = 2b+\beta_0^T\Sigma^{-1}\beta_0+y^Ty-(\beta^*)^T(\Sigma^*)^{-1}\beta^* =$
  $= 2b+\beta_0^T\Sigma^{-1}\beta_0+y^Ty-$
  $((\Sigma^{-1}+X^TX)^{-1}(\Sigma^{-1}\beta_0+X^Ty))^T((\Sigma^{-1}+X^TX)^{-1})^{-1}(\Sigma^{-1}+X^TX)^{-1}\cdot$
  $(\Sigma^{-1}\beta_0+X^Ty) =$
  $= 2b+\beta_0^T\Sigma^{-1}\beta_0+y^Ty-$
  $((\Sigma^{-1}+X^TX)^{-1}(\Sigma^{-1}\beta_0+X^Ty))^T(\Sigma^{-1}\beta_0+X^Ty) =$
  $= 2b+\beta_0^T\Sigma^{-1}\beta_0+y^Ty-$
  $(\Sigma^{-1}\beta_0+X^Ty)^T(\Sigma^{-1}+X^TX)^{-1}(\Sigma^{-1}\beta_0+X^Ty) =$
  $= 2b+\beta_0^T\Sigma^{-1}\beta_0+y^Ty-($
  $\beta_0^T\Sigma^{-1}(\Sigma^{-1}+X^TX)^{-1}\Sigma^{-1}\beta_0+$
  $+2\Sigma^{-1}\beta_0(\Sigma^{-1}+X^TX)^{-1}X^Ty+y^TX(\Sigma^{-1}+X^TX)^{-1}X^Ty) =$
  $= 2b+\beta_0^T\Sigma^{-1}\beta_0+y^Ty-\beta_0^T\Sigma^{-1}\Sigma^*\Sigma^{-1}\beta_0-2\Sigma^{-1}\beta_0\Sigma^*X^Ty-y^TX\Sigma^*X^Ty$
- $(\beta-\beta^*)^T(\Sigma^*)^{-1}(\beta-\beta^*)+2b^* =$
  $= \beta^T\Sigma^{-1}\beta+\beta^TX^TX\beta-2\beta^T\Sigma^{-1}\beta_0-2\beta^TX^Ty+$
  $+\beta_0^T\Sigma^{-1}\Sigma^*\Sigma^{-1}\beta_0+2y^TX\Sigma^*\Sigma^{-1}\beta_0+y^TX\Sigma^*X^Ty+$
  $2b+\beta_0^T\Sigma^{-1}\beta_0+y^Ty-\beta_0^T\Sigma^{-1}\Sigma^*\Sigma^{-1}\beta_0-2\Sigma^{-1}\beta_0\Sigma^*X^Ty-y^TX\Sigma^*X^Ty =$
  $= \beta^T\Sigma^{-1}\beta+\beta^TX^TX\beta-2\beta^T\Sigma^{-1}\beta_0-2\beta^TX^Ty+$
  $2b+\beta_0^T\Sigma^{-1}\beta_0+y^Ty =$
  $= y^Ty-2y^TX\beta+\beta^TX^TX\beta+\beta^T\Sigma^{-1}\beta-2\beta_0^T\Sigma^{-1}\beta+\beta_0^T\Sigma^{-1}\beta_0+2b$ □

Definice 9 (Multivariate student)**.**

$$\mathrm{St}(x|\mu,\Lambda,v) = \frac{\Gamma(|x|/2+v/2)}{\Gamma(v/2)}\frac{|\Lambda|^{\frac{1}{2}}}{(\pi v)^{|x|/2}}\left[1+\frac{((x-\mu)^T\Lambda(x-\mu))^2}{v}\right]^{-|x|/2-v/2}$$

Lemma 18.

$$\int \mathcal{NG}\left(\lambda, w|\mu,\Sigma,a,b\right)d(\lambda,w) = \int \mathrm{St}\left(w|\mu_0, l\left(\Sigma^{-1}\right), v\right)$$

kde $l = a/b, v = 2a$

Důkaz.

- $\int \mathcal{NG}\left(\lambda, w|\mu_0, \Sigma_0, a, b\right) d(\lambda, w) =$
  $= \int \int (2\pi)^{-\frac{|w|}{2}} \left|\Sigma_0\right|^{-\frac{1}{2}} \lambda^{\frac{|w|}{2}} \exp\left(-\frac{1}{2}\lambda\left(\left(w-\mu_0\right)^T \Sigma_0^{-1}\left(w-\mu_0\right)+2b\right)\right) \frac{1}{\Gamma(a)} b^a \lambda^{a-1} d\lambda dw =$
  $= \int (2\pi)^{-\frac{|w|}{2}} \left|\Sigma_0\right|^{-\frac{1}{2}} \frac{1}{\Gamma(a)} b^a \int \lambda^{\frac{|w|}{2}} \exp\left(-\frac{1}{2}\lambda\left(\left(w-\mu_0\right)^T \Sigma_0^{-1}\left(w-\mu_0\right)+2b\right)\right) \lambda^{a-1} d\lambda dw$
- označím $\Delta = \frac{1}{2}\left(\left(w-\mu_0\right)^T \Sigma_0^{-1}\left(w-\mu_0\right)+2b\right), z = \lambda\Delta$
  - $\int \lambda^{\frac{|w|}{2}+a-1} \exp\left(-\frac{1}{2}\lambda\left(\left(w-\mu_0\right)^T \Sigma_0^{-1}\left(w-\mu_0\right)+2b\right)\right) d\lambda =$
    $= \int \lambda^{\frac{|w|}{2}+a-1} \exp(-z) d\lambda = \int z^{\frac{|w|}{2}+a-1} \Delta^{-\left(\frac{|w|}{2}+a-1\right)} \exp(-z) d\lambda =$
    $= \Delta^{-\left(\frac{|w|}{2}+a-1\right)} \int z^{\frac{|w|}{2}+a-1} \exp(-z) d\lambda = \Delta^{-\left(\frac{|w|}{2}+a-1\right)} \int z^{\frac{|w|}{2}+a-1} \exp(-z) \Delta^{-1} dz =$
    $= \Delta^{-\left(\frac{|w|}{2}+a\right)} \int z^{\frac{|w|}{2}+a-1} \exp(-z) dz = \Delta^{-\left(\frac{|w|}{2}+a\right)} \Gamma\left(\frac{|w|}{2}+a\right)$
- $= \int (2\pi)^{-\frac{|w|}{2}} \left|\Sigma_0\right|^{-\frac{1}{2}} \frac{1}{\Gamma(a)} b^a \Delta^{-\left(\frac{|w|}{2}+a\right)} \Gamma\left(\frac{|w|}{2}+a\right) dw$
  $= \int (2\pi)^{-\frac{|w|}{2}} \left|\Sigma_0\right|^{-\frac{1}{2}} \frac{1}{\Gamma(v/2)} \left(\frac{v}{2l}\right)^{v/2} \left(\frac{v}{2l}+\frac{1}{2}\left(\left(w-\mu_0\right)^T \Sigma_0^{-1}\left(w-\mu_0\right)\right)\right)^{-\left(\frac{|w|}{2}+v/2\right)} \cdot$
  $\cdot \Gamma\left(\frac{|w|+v}{2}\right) dw =$
  $= \int (2\pi)^{-\frac{|w|}{2}} \left|\Sigma_0\right|^{-\frac{1}{2}} \frac{1}{\Gamma(v/2)} \left(\frac{v}{2l}\right)^{v/2} \left(\frac{v}{2l}\right)^{-\left(\frac{|w|}{2}+v/2\right)} \cdot$
  $\cdot \left(1+\frac{l}{v}\left(\left(w-\mu_0\right)^T \Sigma_0^{-1}\left(w-\mu_0\right)\right)\right)^{-\left(\frac{|w|}{2}+v/2\right)} \Gamma\left(\frac{|w|+v}{2}\right) dw =$
  $= \int (2\pi)^{-\frac{|w|}{2}} \left|\Sigma_0\right|^{-\frac{1}{2}} \frac{1}{\Gamma(v/2)} \left(\frac{v}{2l}\right)^{-\frac{|w|}{2}} \cdot$
  $\cdot \left(1+\frac{l}{v}\left(\left(w-\mu_0\right)^T \Sigma_0^{-1}\left(w-\mu_0\right)\right)\right)^{-\left(\frac{|w|}{2}+v/2\right)} \Gamma\left(\frac{|w|+v}{2}\right) dw =$
  $= \int \frac{\Gamma\left(\frac{|w|+v}{2}\right)}{\Gamma(v/2)} \frac{\left|l^{-1}\Sigma_0\right|^{-\frac{1}{2}}}{(v\pi)^{1/2}} \left(\frac{v}{2l}\right)^{-\frac{|w|}{2}} \cdot$
  $\cdot \left(1+\frac{1}{v}\left(\left(w-\mu_0\right)^T \left(l^{-1}\Sigma_0\right)^{-1}\left(w-\mu_0\right)\right)\right)^{-\left(\frac{|w|}{2}+v/2\right)} dw = \int \mathrm{St}\left(w|\mu_0, l\left(\Sigma_0^{-1}\right), v\right) dw$

□

2.2 Definice problému

Definice 10 (Model regrese se známou σ^2)**.**

- Jedná se přesně o stejnou defici, jako Definice 1, jen platí, že **neznáme** σ^2 (ale pořád předpokládáme, že nějaká existuje) a **změníme** priorní distribuci na w
- Dále si dáme na σ^2 a w prior tak, že $(w, \lambda) \sim \mathcal{NG}(\mu_0, \Sigma_0, a, b)$ a $\lambda = \sigma^{-2}$, kde dáme konstanty μ_0, Σ_0, a, b
- Tj. konstanty jsou $\mu_0, \Sigma_0, a, b, \{\phi_i\}$, chápu to opět vše jako součást modelu R.

2.3 Prediktivní distribuce

- Opět. Hlavní, co chceme zjistit, je prediktivní distrubuce. Tj.

  $$p(y_{new}|y_{old}, x_{old}, x_{new}, R),$$

  přičemž R budu opět vynechávat

Lemma 19 (Přepis na marginál)**.**

$$p(y_{new}|y_{old}, x_{old}, x_{new}) = \int p(y_{new}|w, \lambda, x_{new}) p(w, \lambda|y_{old}, x_{old}) d(w, \lambda)$$

Pozn.: Nejsem si jist, zda je takovýto zápis integrálu vůbec legální, doufám, že ano. Pokud ne, tak si v duchu představte vektor (w, λ) jako novou proměnnou.
Důkaz.

- Totéž jako u Lemmatu 11. □

Lemma 20 (Přepis posterioru)**.**

$$p(w, \lambda | y_{old}, x_{old}) \propto p(w, \lambda) p(y_{old} | x_{old}, w, \lambda)$$

Důkaz.

- Totéž, co Lemma 12 □

Lemma 21 (Likelihood)**.**

$$y_{new} | w, \lambda, x_{new} \sim \mathcal{N}(\Phi_{new} w, \lambda^{-1} I),$$

kde I je jednotková matice. Podobně $y_{old} | w, \lambda, x_{old} \sim \mathcal{N}(\Phi_{old} w, \lambda^{-1} I)$

Důkaz.

- Totéž co Lemma 13 □

Lemma 22.

$$\ln(p(y_{old} | w, \lambda, x_{old})) = \lambda^{\frac{|y_{old}|}{2}} \left(-\frac{1}{2} \lambda \left(y_{old} - \Phi_{old} w \right)^T \left(y_{old} - \Phi_{old} w \right) \right) + c,$$

kde λ nebereme jako konstantu

Důkaz.

- Víceméně totéž, co v první odrážce důkazu Lemma 16 □

Lemma 23 (Výpočet posterioru)**.** *$w, \lambda | y_{old}, x_{old} \sim \mathcal{NG}(\hat{w}, \hat{\Sigma}, \hat{a}, \hat{b})$, kde*

$$\begin{aligned}
\hat{w} &= (\Sigma_0^{-1} + \Phi_{old}^T \Phi_{old})^{-1} (\Sigma_0^{-1} \mu_0 + \Phi_{old}^T y_{old}) \\
\hat{\Sigma} &= (\Sigma_0^{-1} + \Phi_{old}^T \Phi_{old})^{-1} \\
\hat{b} &= b + \frac{1}{2} (\mu_0^T \Sigma_0^{-1} \mu_0 + y_{old}^T y_{old} - \hat{w}^T \hat{\Sigma}^{-1} \hat{w}) \\
\hat{a} &= a + \frac{1}{2} |y_{old}|
\end{aligned}$$

Důkaz.

- Podle Lemmatu 20 víme, že $p(w, \lambda | y_{old}, x_{old}) \propto p(w, \lambda) p(y_{old} | x_{old}, w, \lambda)$
- Podle Definice 10 $w, \lambda \sim \mathcal{NG}(\mu_0, \Sigma_0, a, b)$
- $\ln p(w, \lambda) = \ln \left(\lambda^{a + \frac{|w| - 2}{2}} \right) \left(-\frac{1}{2} \lambda \left((w - \mu_0)^T \Sigma_0^{-1} (w - \mu_0) + 2b \right) \right) + c$ podle Lemma 16
- Podle Lemma 22 $\ln(p(y_{old} | w, \lambda, x_{old})) = \lambda^{\frac{|y_{old}|}{2}} \left(-\frac{1}{2} \lambda \left(y_{old} - \Phi_{old} w \right)^T \left(y_{old} - \Phi_{old} w \right) \right) + c$
- $\ln p(w, \lambda | y_{old}, x_{old}) = \ln \left(\lambda^{a + \frac{|w| + |y_{old}| - 2}{2}} \right) \cdot$
  $\cdot \left(-\frac{1}{2} \lambda \left((w - \mu_0)^T \Sigma_0^{-1} (w - \mu_0) + \left(y_{old} - \Phi_{old} w \right)^T \left(y_{old} - \Phi_{old} w \right) + 2b \right) \right) + c$
- Použiji Lemma 17 a jednoduchou úpravu exponentu λ
- $\ln p(w, \lambda | y_{old}, x_{old}) = \ln \left(\lambda^{\hat{a} + \frac{|w| - 2}{2}} \right) \left(-\frac{1}{2} \lambda \left((w - \hat{w})^T \hat{\Sigma}^{-1} (w - \hat{w}) + 2\hat{b} \right) \right) + c$

- Podle Lemmatu 16 jde o dané $\mathcal{NG}$. □

Věta 2 (Predictive nedopočítaný)**.**

$$p(y_{new}|y_{old}, x_{old}, x_{new}) \propto \int \text{St}\left(w|\breve{w}, \breve{l}\left(\breve{\Sigma}^{-1}\right), \breve{v}\right) dw,$$

kde

$$\begin{aligned}
\breve{w} &= (\hat{\Sigma}^{-1} + \Phi_{new}^T \Phi_{new})^{-1}(\hat{\Sigma}^{-1}\hat{w} + \Phi_{new}^T y_{new}) \\
\breve{\Sigma} &= (\hat{\Sigma}^{-1} + \Phi_{new}^T \Phi_{new})^{-1} \\
\breve{b} &= \hat{b} + \frac{1}{2}(\hat{w}^T \hat{\Sigma}_0^{-1}\hat{w} + y_{new}^T y_{new} - \breve{w}^T \breve{\Sigma}^{-1}\breve{w}) \\
\breve{a} &= \hat{a} + \frac{1}{2}|y_{new}| \\
\breve{l} &= \breve{a}/\breve{b} \\
\breve{v} &= 2\breve{a}
\end{aligned}$$

Nedokázal jsem, bohužel, dopočítat tenhle integrál do konce.

Bishop uvádí, že konečný integrál je studentovo rozdělení, ale neuvádí výpočet a nechává ho jako cvičení. Je možné, že mám někde ve výpočtu chybu - at' numerickou nebo logickou; nebo jsem Bishopa špatně pochopil; nebo jenom neumím integrovat studentovo rozdělení.

Proto to takhle nechám, protože nevím, co s tím dál. Ještě důkaz.

Důkaz.

- Podle Lemmatu 19 víme, že $p(y_{new}|y_{old}, x_{old}, x_{new}) = \int p(y_{new}|w, \lambda, x_{new}) p(w, \lambda|y_{old}, x_{old}) d(w, \lambda)$
- Podle Lemmatu 21 víme, že $y_{new}|w, \lambda, x_{new} \sim \mathcal{N}(\Phi_{new}w, \lambda^{-1}I)$
- Podle Lemmatu 23 víme, že $w|y_{old}, x_{old} \sim \mathcal{NG}(\hat{w}, \hat{\Sigma}, \hat{a}, \hat{b})$
- Tedy počítám $\int \mathcal{N}\left(y_{new}|\Phi_{new}w, \lambda^{-1}I\right) \mathcal{NG}\left(w, \lambda|\hat{w}, \hat{\Sigma}, \hat{a}, \hat{b}\right) d(w, \lambda) =$
  $= \int \mathcal{N}\left(y_{new}|\Phi_{new}w, \lambda^{-1}I\right) \mathcal{N}\left(w|\hat{w}, \lambda^{-1}\hat{\Sigma}\right) \text{Gam}\left(\lambda|\hat{a}, \hat{b}\right) d(w, \lambda) =$
  $= \int 2\pi^{-|y_{new}|/2} \left|\lambda^{-1}I\right|^{-1} \exp\left(-\frac{1}{2}\lambda\left(y_{new} - \Phi_{new}w\right)^T\left(y_{new} - \Phi_{new}w\right)\right) \cdot$
  $\cdot 2\pi^{-|w|/2} \left|\lambda^{-1}\hat{\Sigma}\right|^{-1} \exp\left(-\frac{1}{2}\lambda(w - \hat{w})^T\hat{\Sigma}\left(w - \hat{w}\right)\right) \lambda^{\hat{a}-1} \exp\left(-\hat{b}\lambda\right) d(w, \lambda) =$
  $= 2\pi^{-|y_{new}|/2} 2\pi^{-|w|/2} \frac{1}{\Gamma(\hat{a})} \hat{b}^{\hat{a}} \cdot$
  $\int \lambda^{|y_{new}|/2 + |w|/2 + \hat{a} - 1} \cdot$
  $\exp\left(-\frac{1}{2}\lambda\left(\left(y_{new} - \Phi_{new}w\right)^T\left(y_{new} - \Phi_{new}w\right) + \left(w - \hat{w}\right)^T\hat{\Sigma}\left(w - \hat{w}\right) + 2\hat{b}\right)\right) d(w, \lambda) \propto$
  $\propto \int \lambda^{|y_{new}|/2 + |w|/2 + \hat{a} - 1} \cdot$
  $\exp\left(-\frac{1}{2}\lambda\left(\left(y_{new} - \Phi_{new}w\right)^T\left(y_{new} - \Phi_{new}w\right) + \left(w - \hat{w}\right)^T\hat{\Sigma}\left(w - \hat{w}\right) + 2\hat{b}\right)\right) d(w, \lambda)$
- použijeme Lemma 17
- $... = \int \lambda^{\breve{a} + \frac{|w|}{2}} \exp\left(-\frac{1}{2}\lambda\left(\left(w - \breve{w}\right)^T \breve{\Sigma}^{-1}\left(w - \breve{w}\right) + 2\breve{b}\right)\right) d(w, \lambda) =$
  $= \int \mathcal{NG}\left(\lambda, w|\breve{w}, \breve{\Sigma}, \breve{a}, \breve{b}\right) d(w, \lambda)$
- Použijeme Lemma 18.

□

3 Zdroje

- Bishop, Christopher M., and Nasser M. Nasrabadi. Pattern recognition and machine learning. Vol. 1. New York: springer, 2006.

– The Bayesian Linear Model with unknown Variance, Simon Kunz http://www.biostat.uzh.ch/teaching/master/previous/seminarbayes/SimonKunz_article.pdf.

– články na anglické wikipedii