Привет !!

ГР.

1. $N(?, !)$
2. ГР
3. ядра?
4. байес. оптимизацию с ГР.

$$x = \begin{pmatrix} x_a \\ x_b \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix}; \Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix}\right)$$

$$x = \begin{pmatrix} x_a \\ x_b \end{pmatrix} \begin{matrix} \}\ a \text{ штук} \\ \}\ b \text{ штук} \end{matrix} \qquad \mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix} \quad [(a+b) \times 1]$$

$$f(x_a, x_b) = \frac{1}{\sqrt{\det(2\pi\Sigma)}} \exp\left(-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)\right)$$

задача: по обуч. выборке строить прогноз для тестовой.

$\updownarrow$

найти $f(x_a | x_b) \sim ?\ N(?; ?)$

$\boxed{\text{Идея 1}}$ $f(x_a | x_b) = \frac{f(x_a, x_b)}{f(x_b)} = \frac{\checkmark}{\checkmark} =$

$x_b \sim N(\mu_b; \Sigma_{bb})$

= (доупрощать до ст-го вида).

Идея 2

2.1. $(x_a | x_b) \sim \boxed{N}(?;\ ?)$

$$= \frac{f(x_a, x_b)}{f(x_b)} = c \cdot \exp\left(\begin{matrix} x_a^T Q x_a + \\ + L x_a \end{matrix} \right)$$

посм-м как на ф-цию от x_a

2.2.

$$x = \begin{pmatrix} x_a \\ x_b \end{pmatrix}$$

$$\operatorname{Cov}(x_a, x_b) = 0 \iff x_a, x_b \text{ незав.}$$

$\Leftarrow$ (для всех распр)

$$f(x_a, x_b) = \text{const} \cdot \exp\left[-\frac{1}{2} \begin{pmatrix} x_a - \mu_a \\ x_b - \mu_b \end{pmatrix}^T \cdot \begin{pmatrix} \Sigma_{aa} & 0 \\ 0 & \Sigma_{bb} \end{pmatrix}^{-1} \cdot \begin{pmatrix} x_a - \mu_a \\ x_b - \mu_b \end{pmatrix} \right]$$

диаг.

$$= \text{const} \exp\left(-\frac{1}{2} (x_a - \mu_a)^T \cdot \Sigma_{aa}^{-1} (x_a - \mu_a) \right) \cdot$$
$$\cdot \exp\left(-\frac{1}{2} (x_b - \mu_b)^T \cdot \Sigma_{bb}^{-1} (x_b - \mu_b) \right)$$

2.3.

Упр.

$$x_a = c + D\cdot x_b + \varepsilon$$

c? D? $\begin{cases} E(\varepsilon) = 0 \\ \operatorname{Cov}(x_b, \varepsilon) = 0. \end{cases}$

$E(LHS) = E(RHS)$

$\operatorname{Cov}(LHS, x_b) = \operatorname{Cov}(RHS, x_b)$

$$\begin{cases} E(x_a) = c + D\cdot E(x_b) + 0 \\ \operatorname{Cov}(x_a, x_b) = 0 + D\cdot \operatorname{Cov}(x_b, x_b) + \underbrace{0}_{\operatorname{Cov}(\varepsilon, x_b)} \end{cases}$$

$$\begin{cases} \mu_a = c + D\cdot \mu_b \\ \Sigma_{ab} = D\cdot \Sigma_{bb} \end{cases} \qquad \begin{cases} D = \Sigma_{ab}\cdot \Sigma_{bb}^{-1} \\ c = \mu_a - \Sigma_{ab}\Sigma_{bb}^{-1}\mu_b \end{cases}$$

$$x_a = c + D\cdot x_b + \varepsilon$$

$$E(x_a | x_b) = E(c + D x_b + \varepsilon | x_b) =$$

$$= c + D\cdot x_b + \underbrace{E(\varepsilon | x_b)}_{= 0} =$$

$$= \mu_a + (\Sigma_{ab}\cdot\Sigma_{bb}^{-1})\cdot(x_b - \mu_b)$$

$$\operatorname{Var}(x_a | x_b) = \operatorname{Var}(c + D x_b + \varepsilon | x_b) =$$

$$= \operatorname{Var}(\varepsilon | x_b) = \operatorname{Var}(\varepsilon) \overset{?}{=} \Sigma_{aa} - \Sigma_{ab}\cdot\Sigma_{bb}^{-1}\cdot\Sigma_{bb}\cdot\Sigma_{bb}^{-1}\cdot\Sigma_{ba}$$

$$\operatorname{Var}(LHS) = \operatorname{Var}(RHS)$$

$$\operatorname{Var}(x_a) = \operatorname{Var}(c + D\cdot x_b + \varepsilon) = D\cdot\underbrace{\operatorname{Var}(x_b)}_{\Sigma_{bb}}\cdot D^T + \operatorname{Var}(\varepsilon)$$

$(x_a|x_b) \sim$

$\sim N\left(\mu_a + \Sigma_{ab}\cdot\Sigma_{bb}^{-1}(x_b-\mu_b);\ \Sigma_{aa}-\Sigma_{ab}\cdot\Sigma_{bb}^{-1}\cdot\Sigma_{ba}\right)$

$\boxed{\text{Идея 3}}$ посмотреть на d, $d^2 \ln f$

Σ - ков. м-ца

$u \sim N(\mu;\Sigma)$ $\qquad \Sigma^{-1} = V$ - м-ца точности

$$\ln f = \ln c - \frac{1}{2}(u-\mu)^T\cdot\Sigma^{-1}\cdot(u-\mu) =$$

$$= \ln c - \frac{1}{2}(u-\mu)^T\cdot V\cdot(u-\mu)$$

$$d\ln f = -(u-\mu)^T\cdot V\cdot du = -du^T\cdot V\cdot(u-\mu)$$

$$d\ln f = 0 \iff \boxed{u^* = \mu}$$

$$d^2\ln f = \boxed{-du^T\cdot V\cdot du} = -du^T\cdot\Sigma^{-1}\cdot du$$

как ф-ция от x_a

$$f(x_a|x_b) = \frac{f(x_a, x_b)}{f(x_b)} = c_1\cdot f(x_a,x_b) \qquad \begin{bmatrix} V_{aa} & V_{ab} \\ V_{ba} & V_{bb}\end{bmatrix}$$

$$\ln f(x_a|x_b) = c_2 - \frac{1}{2}\left[\begin{pmatrix} x_a-\mu_a \\ x_b-\mu_b\end{pmatrix}^T\cdot V\cdot\begin{pmatrix} x_a-\mu_a \\ x_b-\mu_b\end{pmatrix}\right] =$$

$$= c_3 - \frac{1}{2}(x_a-\mu_a)^T\cdot V_{aa}\cdot(x_a-\mu_a) - (x_a-\mu_a)\cdot V_{ab}\cdot(x_b-\mu_b)$$

$$d\ln f(x_a|x_b) = -dx_a^T\cdot\left[V_{aa}(x_a-\mu_a)+V_{ab}\cdot(x_b-\mu_b)\right]$$

$$\cdot \left[V_{aa}(x_a^* - \mu_a) + V_{ab} \cdot (x_b - \mu_b) \right] = 0$$

при $\quad x_a^* = E(x_a | x_b)$

$$E(x_a|x_b) = x_a^* = \mu_a - V_{aa}^{-1} \cdot V_{ab} \cdot (x_b - \mu_b)$$

$$\operatorname{Var}(x_a | x_b) = \qquad d^2 \ln f(x_a|x_b) =$$

$$= - dx_a^T \cdot V_{aa} \cdot dx_a$$

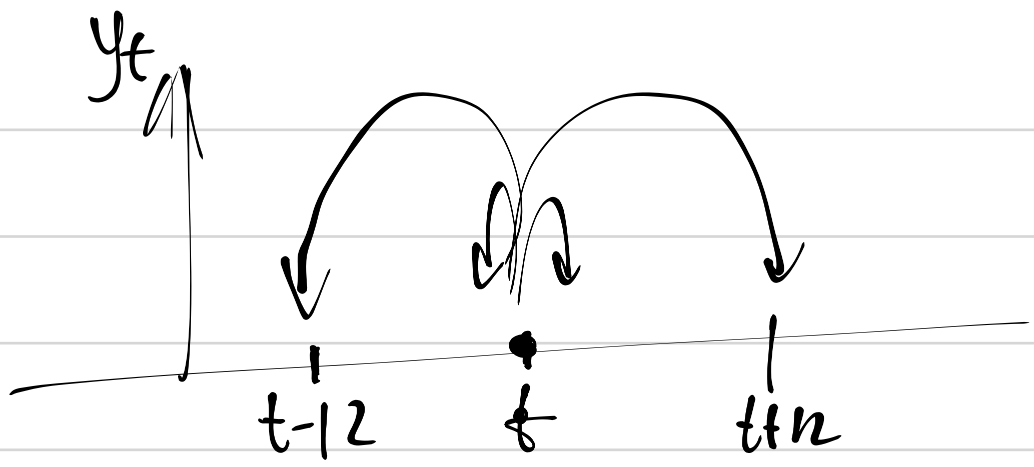
$$= V_{aa}^{-1}$$

$$(x_a|x_b) \sim$$

$$\sim N\Big(\mu_a + \Sigma_{ab} \cdot \Sigma_{bb}^{-1} (x_b - \mu_b);\ \underbrace{\Sigma_{aa}}_{\operatorname{Var}(x_a)} - \underbrace{\Sigma_{ab} \cdot \Sigma_{bb}^{-1} \cdot \Sigma_{ba}}_{\operatorname{Var}(z)}\Big)$$

$$\sim N\left(\mu_a - V_{aa}^{-1} V_{ab} (x_b - \mu_b);\ V_{aa}^{-1}\right)$$

$$\Sigma = \begin{pmatrix} \Sigma_{aa} & \Sigma_{ab} \\ \Sigma_{ba} & \Sigma_{bb} \end{pmatrix} \qquad \Sigma^{-1} = V = \begin{pmatrix} V_{aa} & V_{ab} \\ V_{ba} & V_{bb} \end{pmatrix}$$



$$x_1, \dots, x_n \qquad \begin{pmatrix} f(x_1) \\ \vdots \\ f(x_n) \end{pmatrix} \sim N\left(\begin{pmatrix} m(x_1) \\ \vdots \\ m(x_n) \end{pmatrix} ; \begin{bmatrix} K(x_1,x_1) & K(x_1,x_n) \\ \dots & K(x_n,x_n) \end{bmatrix} \right)$$

$$\boxed{f(x) \sim GP(m, K)}$$

K - ядерная ф-ция – скалярное произв-е в пр-нном пр-ве.

$$K(x_L, x_R) = \langle \varphi(x_L), \varphi(x_R) \rangle$$

LIN: $\alpha + \beta \cdot x_L \cdot x_R$

RBF: $\alpha \cdot \exp(-\beta \cdot (x_L - x_R)^2)$

PER: $\alpha \cdot \exp\left(-\beta \sin^2\left((x_L - x_R) \cdot \frac{\pi}{T}\right)\right)$

COS: $\alpha \cos(\beta(x_L - x_R))$

...

способы создания новых ядер из старых:

$K_1 + K_2 \to K_{new}$

$K_1 \cdot K_2 \to K_{new}$

$K_{old}(\varphi(x_L), \varphi(x_R)) \to K_{new}$

$\alpha K_{old} \to K_{new}$

$\langle \varphi(x_L), \varphi(x_R) \rangle \to K_{new}$

Модель:

в рядах: x_i – номер набл.

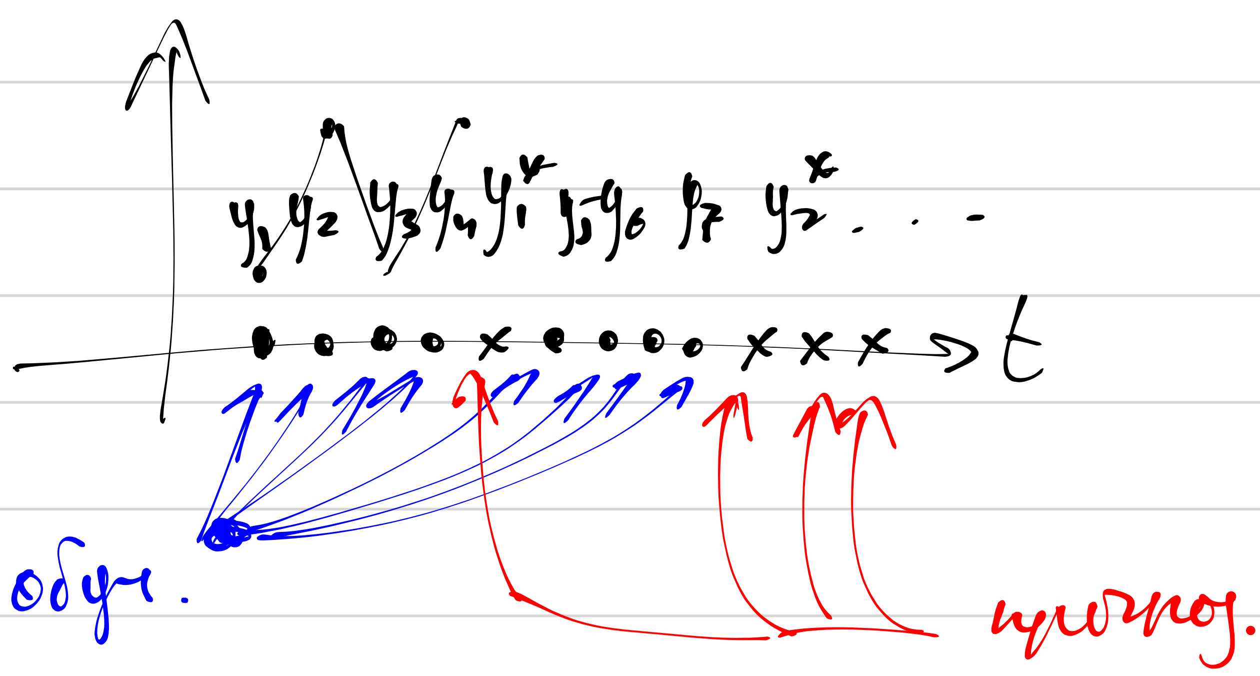
$$y_i = f(x_i) + u_i$$

$$y_* = f(x_*) + u_*$$

$$u_i \sim N(0; \sigma^2)$$

$$f \sim GP(m, k)$$

незав.



$$\begin{pmatrix} y \\ y^* \end{pmatrix} \sim N\left(\begin{pmatrix} m(x) \\ m(x^*) \end{pmatrix}; \begin{bmatrix} k(x,x)+\sigma^2 I & k(x,x^*) \\ k(x_*,x) & k(x_*,x_*)+\sigma^2 I \end{bmatrix}\right)$$

упр.

$$(y^* \mid y) \sim N(\_\_\_\_\_\_ ; \_\_\_\_\_\_)$$

искусство!

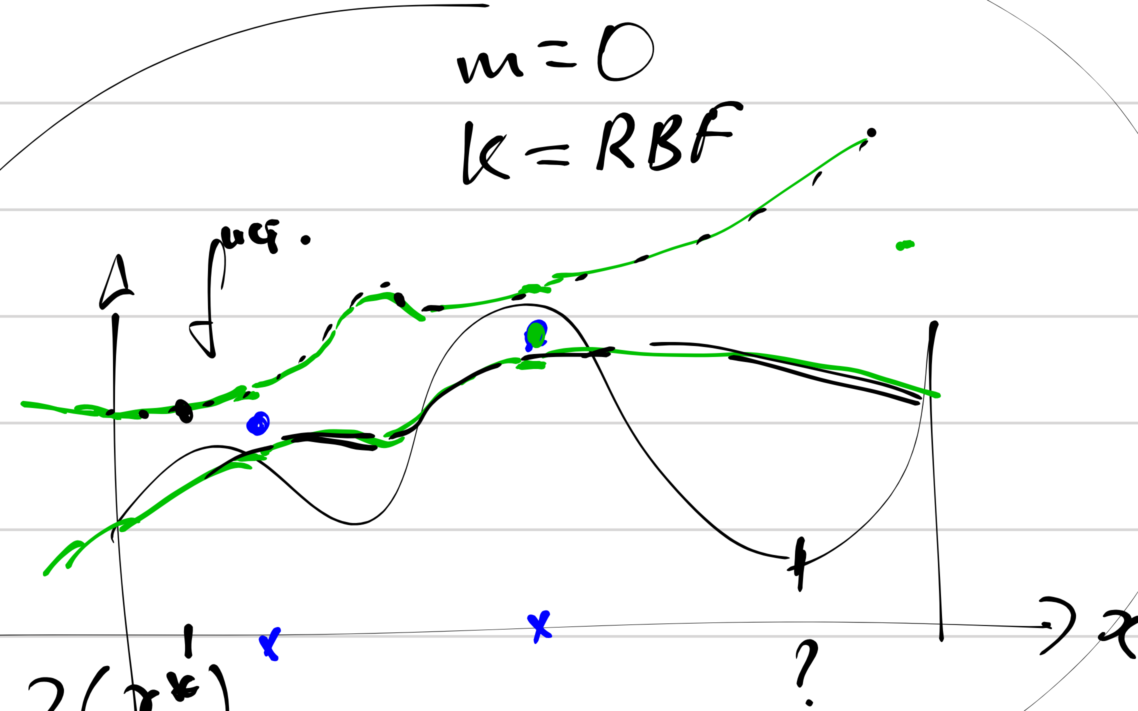
$$k = LIN + PER + RBF_1 \cdot COS_1 + RBF_2 \cdot COS_2$$

Шаг в сторону.

Байесовская опт-ция с GP.

$$\begin{cases} y_i = f(x_i) + u_i \\ \boxed{f \sim GP(m, \underline{k})} \; \text{ядро} \\ u_i \sim N(0; \sigma^2) \end{cases}$$

$\boxed{\min f \;?}$



$$LCB = \mu(x^*) - k \cdot \sigma(x^*)$$

$$PofI(x^*) = P\left(f(x^*) < y_{best}\right) =$$

$$= P\left(\frac{f(x^*) - \mu(x^*)}{\sigma(x^*)} < \frac{y_{best} - \mu(x^*)}{\sigma(x^*)}\right) =$$

$$= \Phi\left(\frac{y_{best} - \mu(x^*)}{\sigma(x^*)}\right) \to \max_{x^*}$$

$$\mu(x^*) = E\left(f(x^*) \mid y\right)$$

$$\sigma^2(x^*) = Var\left(f(x^*) \mid y\right)$$

$$EI(x^*) = E\left(\min\left(f(x^*) - y_{best}, 0\right)\right) =$$

$$= -\sigma(x^*) \cdot \left[\varphi\left(\frac{y_b - \mu}{\sigma}\right) + \frac{y_b - \mu}{\sigma}\Phi\left(\frac{y_b - \mu}{\sigma}\right)\right].$$