$$p(x|z)\,p(z) = p(x,z)$$

$$p(z|x) = \frac{p(x|z)\,p(z)}{\int_z p(x|z)\,p(z)\,dz} \approx$$

$$\approx \delta(z - z_{MP}) = \underset{q \in \Delta}{\arg\min}\; KL\left(q(z)\,\|\,p(z|x)\right)$$

$$p(x|z_1, z_2)\,p(z_1, z_2) = p(x, z_1, z_2)$$ условная сопряжённость

$$p(z_1, z_2|x) \approx q(z_1)\cdot q(z_2)$$

$$p(z_1, z_2|x) \approx q(z_1)\cdot \delta(z_2 - z_{MP}^2)$$

$$p(z_1, z_2|x) \approx \delta(z_1 - z_{MP}^1)\,\delta(z_2 - z_{MP}^2)$$

$$\ln q(z_i) \overset{const}{=} \mathbb{E}_{q(z_{\setminus i})} \ln p(x,z)$$

$$q(z_i) = \frac{\exp\left(\mathbb{E}_{z_{\setminus i}} \ln p(x,z)\right)}{\int -\!\!\|\!\!-\, dz_i} = \delta(z_i - z_{MP}^i)$$

$$z_{MP}^i = \underset{z_i}{\arg\max}\; \mathbb{E}_{z_{\setminus i}} \ln p(x,z)$$

$$p(\theta)\,p(x,z|\theta) = p(x|z,\theta)\,p(z|\theta)\,p(\theta) = p(x,z,\theta)$$

$$EM':\; \theta_{MP} = \underset{\theta}{\arg\max}\; p(\theta|x)$$

$$E\text{-step}:\; q(z) = p(z|x,\theta) = \frac{p(x,z,\theta)}{\int p(x,z,\theta)\,dz} =$$

$$= \frac{p(x|z,\theta)\,p(z|\theta)\,\cancel{p(\theta)}}{\int p(x|z,\theta)\,p(z|\theta)\,\cancel{p(\theta)}\,dz} = \frac{p(x|z,\theta)\,p(z|\theta)}{\int p(x|z,\theta)\,p(z|\theta)\,dz}$$

M - step:

$$\mathbb{E}_{q(z)} \ln p(x, z, \theta) \to \max$$

$$\mathbb{E}_{q(z)} \ln p(x, z|\theta) + \ln p(\theta) \to \max$$

// $\ln p(\theta|x) = \ln p(x|\theta) + \ln p(\theta) + \text{const}$

Св-ва ВМ	Метод Вывода	Приближения апостериорного распределения
FM conjugacy	Th Bayes	$p(z, \theta \| x)$
Conditional conjugacy on z and θ	Mean-Field	$q(z)q(\theta) = \arg\min\limits_{q \in Q} KL(q(z,\theta) \\| p(z,\theta\|x))$
Conjugacy on z given θ fixed	EM'	$q(z)\delta(\theta - \theta_{MP}) = \arg\min\limits_{\substack{q(z)\,\forall \\ q(\theta) \in \Delta}} KL(q(z,\theta) \\| p(z,\theta\|x))$

$$\ln q(\theta) \overset{const}{=} \mathbb{E}_{q(z)} \ln p(x, z, \theta) \quad \{M\ max!\}$$

$$\ln q(z) \overset{const}{=} \mathbb{E}_{q(\theta)} \ln p(x, z, \theta) = \ln p(x, z, \theta_{MP})$$

$$q(z) = \frac{p(x, z, \theta_{MP})}{\int_z p(x, z, \theta_{MP})\, dz} \quad \{E\ max!\}$$

Conjugacy on θ given z fixed	ME'	$q(\theta)\delta(z - z_{MP})$

Conjugacy conditional on $z_1 \dots z_m$ given Θ fixed — Variational EM' {Variational} ME' — $q(z_1)\cdots q(z_m)\,\delta(\Theta-\Theta_{MP})$

No conjugacy — Crisp EM' — $\delta(z-z_{MP})\,\delta(\Theta-\Theta_{MP})$

$(z,\Theta) = \arg\max_{z,\Theta} p(x,z,\Theta)$

Gaussian Process (GP)

$f(x)\ ,\ x \in \mathbb{R}^D$

$\forall n \quad x_1 \dots x_n : \ (f(x_1)\dots f(x_n)) \sim \mathcal{N}\left(f(x_1)\dots f(x_n) \mid (m(x_1)\dots m(x_n)),\ \begin{bmatrix} k(x_1,x_1) \dots k(x_1,x_n) \\ \dots \\ k(x_n,x_1) \dots k(x_n,x_n) \end{bmatrix}\right)$

конечномерная проекция

$m(x) \overset{def}{=} \mathbb{E} f(x)$

$k(x,x') \overset{def}{=} \mathrm{Cov}(f(x), f(x'))$

стационарное распределение

инвариантность относительно сдвига

$(f(x_1+x_0),\dots,f(x_n+x_0)) \sim (f(x_1),\dots,f(x_n))$

стационарность в узком смысле

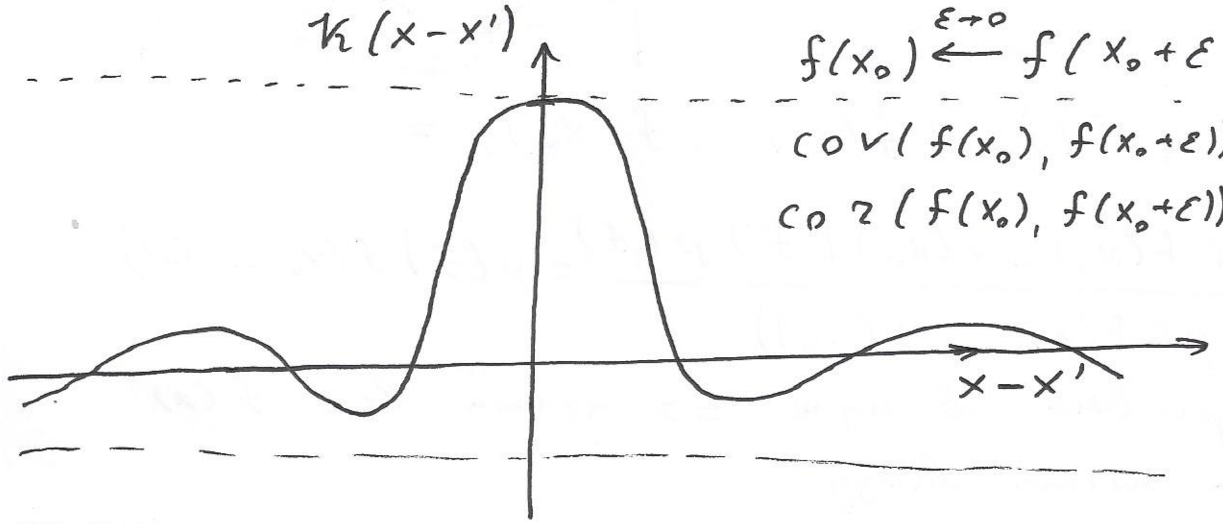
$m(x) = const\ ,\ k(x,x') = \mathcal{K}(x-x')$ ковариационная ф-ия

стационарность в широком смысле

$f(x_0) \xleftarrow{\varepsilon\to 0} f(x_0+\varepsilon)$

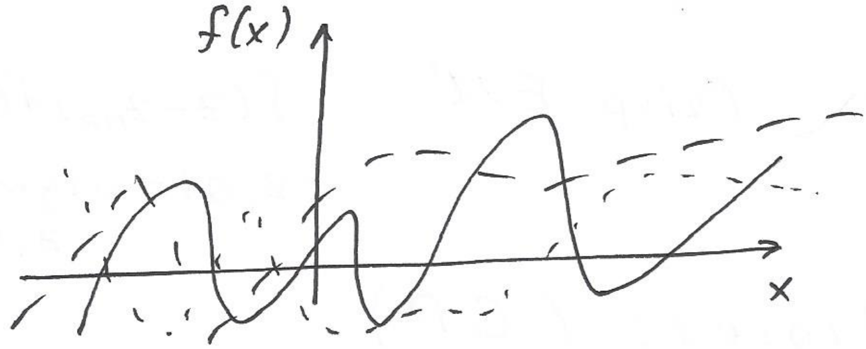
$\mathrm{cov}(f(x_0), f(x_0+\varepsilon)) \xrightarrow{\varepsilon\to 0} \mathcal{K}(0)$

$\mathrm{cor}(f(x_0), f(x_0+\varepsilon)) \xrightarrow{\varepsilon\to 0} 1$



Непрерывность $k(x-x_0)$ в нуле $\Rightarrow$ ~~$f(x)$ непрерывна почти всюду.~~

Почти все реализации $f(x)$ непрерывны почти всюду.



$$p(f(x) \mid f(x_1) \dots f(x_n)) = \frac{p(f(x), f(x_1), \dots, f(x_n))}{p(f(x_1) \dots f(x_n))} = \cancel{N(f(x))}$$

$$= N(f(x) \mid \mu, \sigma^2)$$

формулы Андерсона

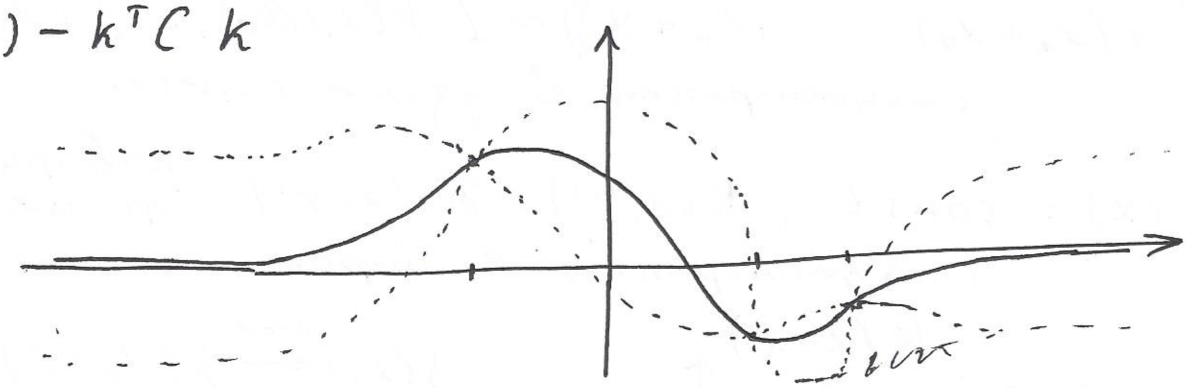
$$\begin{pmatrix} C & k \\ k^T & k(x,x) \end{pmatrix}$$

$$C_{ij} = k(x_i, x_j)$$

$$k_i = k(x, x_i)$$

$$\mu = k^T C^{-1} f$$

$$\sigma^2 = k(0) - k^T C^{-1} k$$



$$p(f(\hat{x}_1) \dots f(\hat{x}_n) \mid f(x_1) \dots f(x_n)) =$$

$$= \frac{p(f(\hat{x}_1) \dots f(x_n) \mid f)\, p(f)}{p(f(x_1) \dots f(x_n))} = p(f \mid f(x_1 \dots x_n))$$

$k(x)$ разрывна в нуле $\Rightarrow$ почти все $f(x)$ разрывны почти всюду

$(X, T) = (x_i, t_i)_{i=1}^{n}$, $x_i \in \mathbb{R}^D$, $t_i \in \mathbb{R}$

Восстановление регрессии — обуславливание гауссовского сл. п.

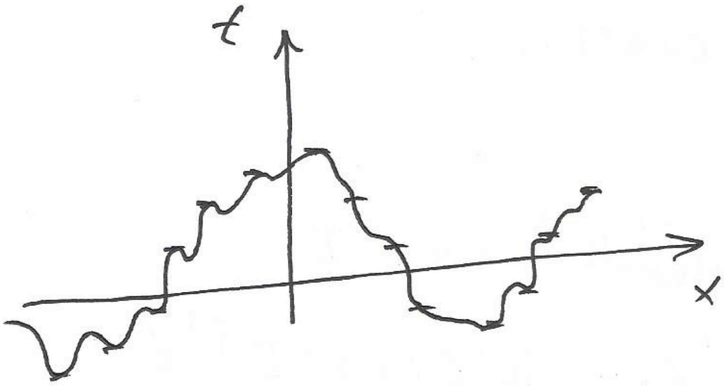
$$p(t_1 \dots t_n \mid f(x_1) \dots f(x_n)) = \mathcal{N}(t_1 \dots t_n \mid (f(x_1) \dots f(x_n)), \sigma^2 I) =$$
$$= p(t_1 \dots t_n \mid f)$$

$$p(f \mid t_1 \dots t_n) = \frac{p(t_1 \dots t_n \mid f)\, p(f \mid K)}{\underbrace{\int p(t_1 \dots t_n \mid f)\, p(f \mid K)\, df}_{\text{Evidence}}}$$

$$\int h(a) p(a, b)\, da\, db = \int h(a) p(a)\, da$$

$$\int p(t_1 \dots t_n \mid f)\, p(f \mid K)\, df = \int p(t_1 \dots t_n \mid f(x_1) \dots f(x_n)) \cdot$$
$$\cdot\, p(f \mid K)\, df = \int p(t_1 \dots t_n \mid f(x_1) \dots f(x_n))\, p(f(x_1) \dots f(x_n) \mid K) \cdot$$
$$\cdot\, df(x_1) \dots df(x_n) = \mathcal{N}(t_1 \dots t_n \mid 0, \tilde{C})$$

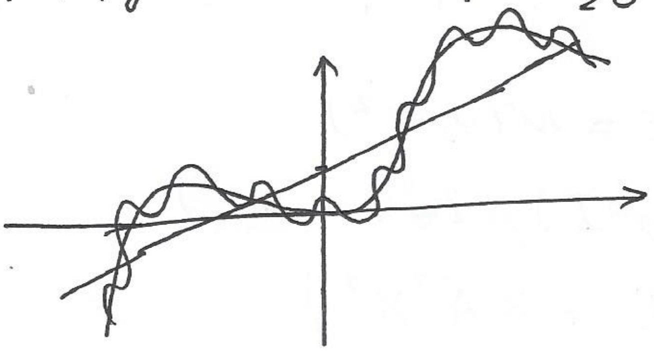
$$\tilde{C}_{ij} = K_\alpha(x_i, x_j) + \sigma^2 I\, [x_i = x_j]$$



$$K_\alpha(x - x') = \alpha_0 e^{-\sum_{j=1}^{D} \alpha_j (x_j - x_j')^2}$$

28.10.16 было сем

$$K(x, y) = A \exp\left(-\frac{\|x - y\|^2}{2\sigma^2}\right) + \sigma^2 [x = y] + a + x^T y$$



$$\begin{cases} f \sim GP(0, k(x,y)) \\ t_n = f(x_n) + \varepsilon, \quad \varepsilon \sim N(0, \sigma^2) \end{cases} \iff \begin{cases} f \sim GP(0, k(x,y) + \sigma^2[x=y]) \\ t_n = f(x_n) \end{cases}$$

$$\{t, x\}^N \quad \{t_{new}, X_{new}\}^M$$

$$p(t_{new} \mid X_{new}, t, x, k(x,y), \sigma^2) =$$

$$= \cancel{p(t_{new}, X_{new}, t} $$

$$= \frac{p(t_{new}, t \mid X_{new}, X, k(x,y), \sigma^2)}{p(t \mid k(x,y), \sigma^2)}$$

$$p(t_{new}, t \mid k(x,y), \sigma^2) = N\left(0, \begin{array}{cc} N \times N & N \times M \\ \left[\begin{array}{c|c} C + \sigma^2 I & k \\ \hline k^T & k_{new} + \sigma^2 I \end{array}\right] & \\ M \times N & M \times M \end{array}\right)$$

$$p(t_{new} \mid X_{new}, t, x, k(x,y), \sigma^2) =$$

$$= N(t_{new} \mid -\Lambda_{bb}^{-1} \Lambda_{ab}^T t, \; \Lambda_{bb}^{-1}) \;\#$$

$$\Lambda_{bb} = (k_{new} + \sigma^2 I, - k^T (C + \sigma^2 I) k)^{-1}$$

$$\Lambda_{ab} = -(C + \sigma^2 I)^{-1} k \Lambda_{bb}$$

$$\mu = -\Lambda_{bb}^{-1} \Lambda_{ab}^T t = \cancel{+\Lambda_{bb}^{-1} \Lambda_{bb}^{-1} k^{-1} (C + \sigma^2 I)}$$

$$= +\Lambda_{bb}^{-1} \Lambda_{bb}^T k^T ((C + \sigma^2 I)^{-T} t = k^T (C + \sigma^2 I)^{-T} t$$

$$k_{new} + \sigma^2 I - k^T (C + \sigma^2 I) k = \Sigma_*$$

RVM
$$\begin{cases} w \sim N(0, A^{-1}) \\ t_n = w^T x_n + \varepsilon, \quad \varepsilon = N(0, \sigma^2) \end{cases}$$

$$p(t \mid w) p(w) = N(t \mid X w, \sigma^2 I) p(w \mid 0, A^{-1})$$

$$p(t \mid x, \sigma^2, A) = N(t \mid 0, \sigma^2 I + X A^{-1} X^T)$$

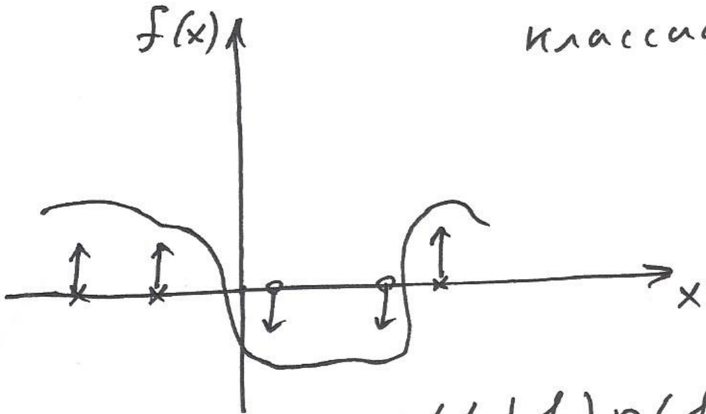
$$K(x,y) = \sigma^2 I[x=y] + x^T A^{-1} y$$

$$\Sigma = \sigma^2 I + X A^{-1} X^T$$

RVM как GP

RBF преобразование

$$x_{new} \quad \begin{matrix} x_1 \\ \vdots \\ x_N \end{matrix} \quad \begin{bmatrix} \exp\left(-\frac{\|x_{new} - x_1\|^2}{\gamma}\right) \\ \dots \\ \exp\left(-\frac{\|x_{new} - x_N\|^2}{\gamma}\right) \end{bmatrix}$$



классификация GP

$t_i \; \{\pm 1\}$

$sgn(f(x_i))$

$$p(t_i \mid f(x_i)) = \frac{1}{1 + \exp(-t_i f(x_i))}$$

$$p(f \mid t) = \frac{p(t \mid f)\, p(f)}{\int p(t \mid f)\, p(f)\, dt} \propto p(t \mid f)\, p(f) =$$

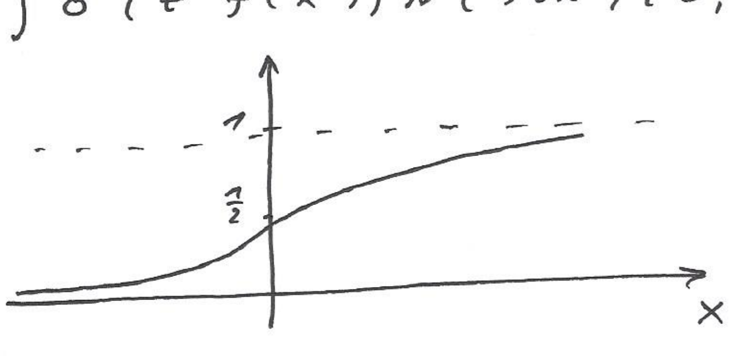
$$= \prod_n \sigma(t_n f(x_n))\, p(f(x_1) \dots f(x_n) \mid K(x,y))$$

обучение? $\Leftrightarrow K(x,y)$?

$1 \longrightarrow \max\limits_{f(x_1) \dots f(x_n)}$ $\quad 2 \longrightarrow \max\limits_{K}$

$$p(t^* \mid x^*, x, t) = \int p(t^* \mid f(x^*))\, p(f(x^*))\, d f(x^*) =$$

$$= \int \sigma(t^* f(x^*))\, N(f(x^*) \mid 0, \rho^2)\, d f(x^*) \ominus$$



$$\Phi(y) = \int_{-\infty}^{y} N(z \mid 0, 1)\, dz$$

$$\sigma(y) \approx \Phi(\lambda y)$$

$$= \int \sigma(y)\, N(y \mid 0, s^2)\, dy = \int_{-\infty}^{\infty} \int_{-\infty}^{\lambda y} N(z \mid 0, 1)\, dz \,\cdot$$

$$\cdot\, N(y \mid 0, s^2)\, dy = \mathbb{P}(\lambda y \geq z) = \mathbb{P}(\eta \overset{<}{\cancel{\geq}} \text{const}) = \Phi(\text{const}) \approx$$

$$\approx \sigma(\text{const}')$$