Homework #1

John Karasev

STAT 672

January 22, 2020

1 Orthogonal projection in a hilbert space

$$\pi_V(f) \in V;$$
$$\langle f - \pi_V(f), g \rangle = 0 \; \forall g \in V;$$

1.
$$\begin{aligned}\|f\|^2 &= \langle f, f\rangle \\ &= \langle \pi_v(f) + (f - \pi_v(f)), \pi_v(f) + (f - \pi_v(f))\rangle \\ &= \langle \pi_v(f), \pi_v(f)\rangle + 2\langle \pi_v(f), f - \pi_v(f)\rangle + \langle f - \pi_v(f), f - \pi_v(f)\rangle \\ &= \|\pi_v(f)\|^2 + \|f - \pi_v(f)\|^2 \; \textit{center term cancels since orthogonal}\end{aligned}$$

2.
$$\begin{aligned}\langle f, \pi_v(g)\rangle &= \langle (f - \pi_v(f)) + \pi_v(f), \pi_v(g)\rangle \\ &= \langle f - \pi_v(f), \pi_v(g)\rangle + \langle \pi_v(f), \pi_v(g)\rangle \\ &= \langle \pi_v(f), \pi_v(g)\rangle \; \textit{by orthogonality}\end{aligned}$$

$$\begin{aligned}\langle g, \pi_v(f)\rangle &= \langle (g - \pi_v(g)) + \pi_v(g), \pi_v(f)\rangle \\ &= \langle g - \pi_v(g), \pi_v(f)\rangle + \langle \pi_v(g), \pi_v(f)\rangle \\ &= \langle \pi_v(g), \pi_v(f)\rangle \; \textit{by orthogonality}\end{aligned}$$

$$\langle g, \pi_v(f)\rangle = \langle f, \pi_v(g)\rangle$$

3.

2 A reproducing kernel is positive definite

1.
$$K(x, y) = \langle K_x(y), K_y(x)\rangle \; \textit{by reproducing property}$$
$$K(y, x) = \langle K_y(x), K_x(y)\rangle \; \textit{by reproducing property}$$
$$K(x, y) = K(y, x)$$

2.
$$\begin{aligned}K(x, y) &= \langle K_x(y), K_y(x)\rangle \\ &= \langle K_y(x), K_y(x)\rangle \geq 0 \; \textit{by symmetry}\end{aligned}$$

3.
$$\begin{aligned}|f(x)| &\leq \|f\|\sqrt{K(x,x)} \\ |\langle f, K_x\rangle| &\leq \|f\|\sqrt{\langle K_x, K_x\rangle} \\ |\langle f, K_x\rangle| &\leq \|f\|\|K_x\| \; \textit{true by Cauchy} - \textit{Schwartz inequality}\end{aligned}$$

3 RKHS over a finite set

1.
$$\langle \boldsymbol{f}, \boldsymbol{g} \rangle = \boldsymbol{f}^T \boldsymbol{K}^{-1} \boldsymbol{g} \quad \boldsymbol{f}, \boldsymbol{g} \in \mathbb{R}^{2m+1}$$

$$\boldsymbol{\mathcal{N}}(\boldsymbol{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{(2\pi)^{\frac{n}{2}} |\boldsymbol{\Sigma}|^{\frac{1}{2}}} exp\left\{ \frac{-(\boldsymbol{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu})}{2} \right\} \quad \boldsymbol{x} \in \mathbb{R}^n$$

$$\begin{aligned} p(\boldsymbol{f}) &= C e^{-\frac{\|\boldsymbol{f}\|_H^2}{2}} \\ &= C exp\left\{ -\frac{\boldsymbol{f}^T \boldsymbol{K}^{-1} \boldsymbol{f}}{2} \right\} \\ &= \frac{1}{(2\pi)^{\frac{2m+1}{2}} |\boldsymbol{K}|^{\frac{1}{2}}} exp\left\{ -\frac{\boldsymbol{f}^T \boldsymbol{K}^{-1} \boldsymbol{f}}{2} \right\} \\ p(\boldsymbol{f}) &= \boldsymbol{\mathcal{N}}(\boldsymbol{f}; \boldsymbol{0}, \boldsymbol{K}) \end{aligned}$$

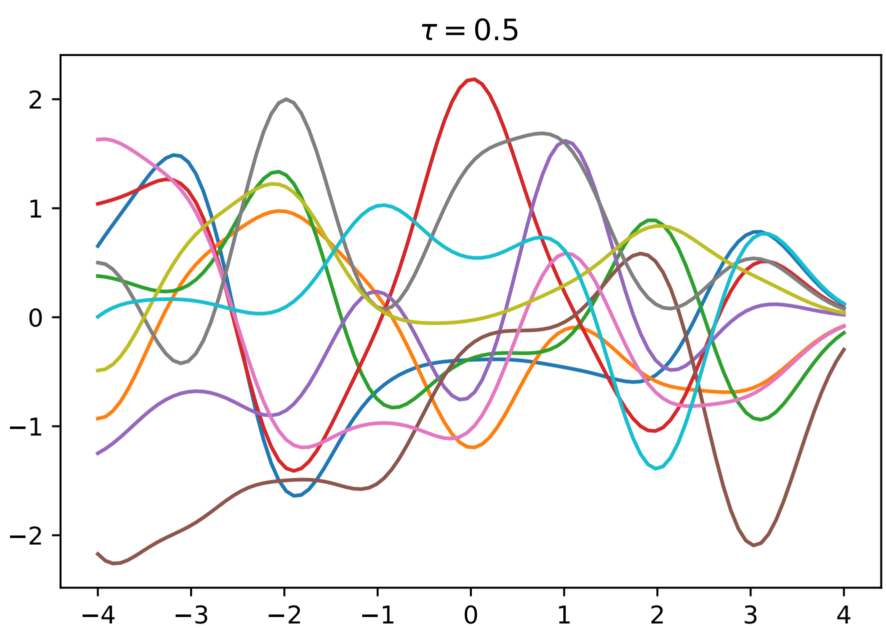
$$\begin{aligned} \therefore \ \boldsymbol{\mu} &= \boldsymbol{0} \\ \boldsymbol{\Sigma} &= \boldsymbol{K} \end{aligned}$$

2. C is a normalizing constant so the probability distribution sums to 1

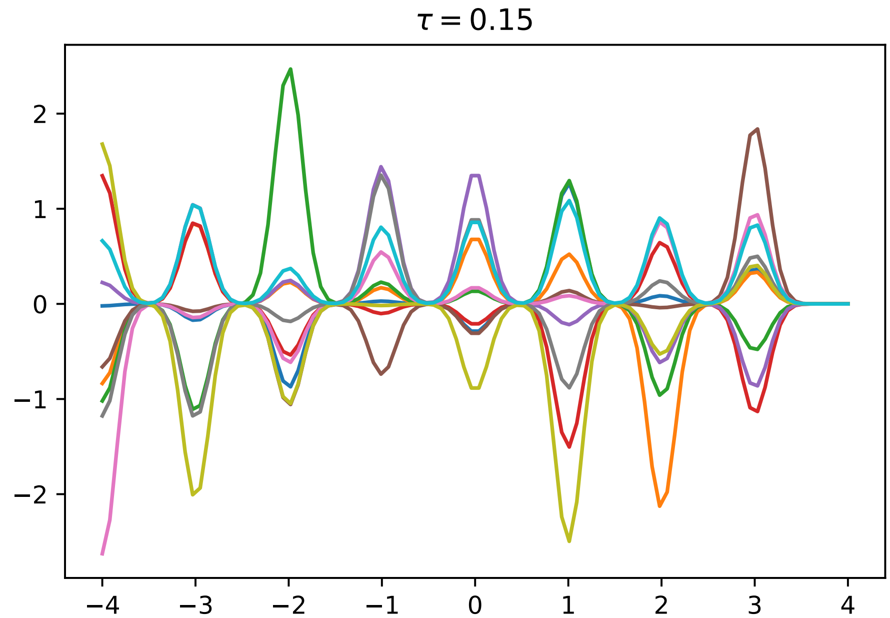
$$C = \frac{1}{(2\pi)^{\frac{2m+1}{2}} |\boldsymbol{K}|^{\frac{1}{2}}}$$

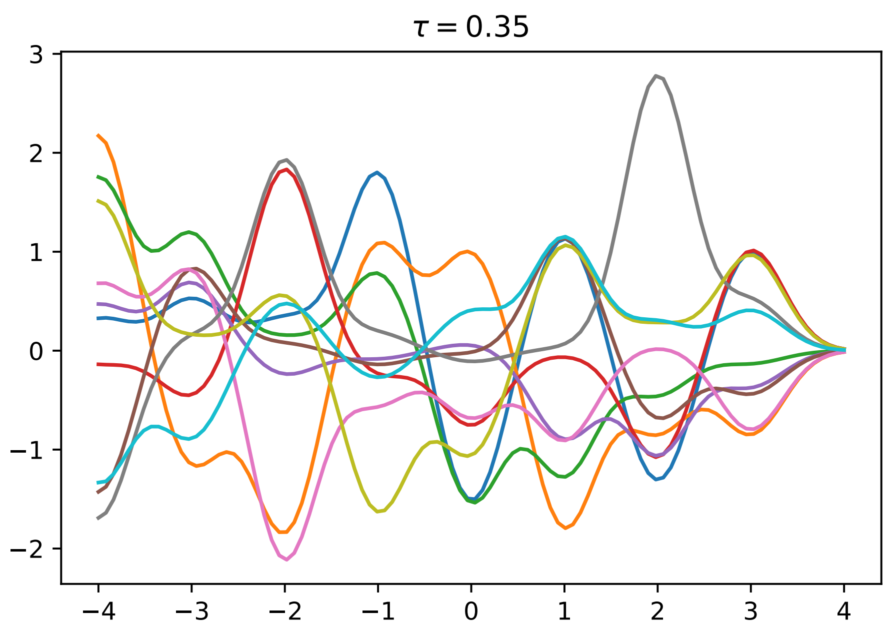
which is taken from the $\boldsymbol{\mathcal{N}}(\boldsymbol{f}; \boldsymbol{0}, \boldsymbol{K})$ normalization constant

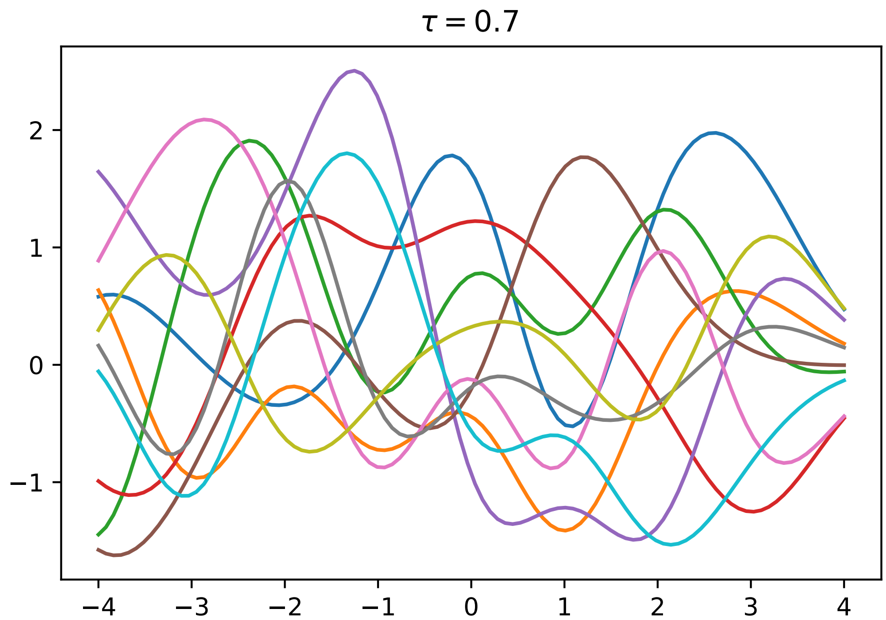
3. (a) kernel matrix: $\boldsymbol{K} = [k_{ij}] = Ce^{-\frac{(i-j)^2}{2\tau^2}}$



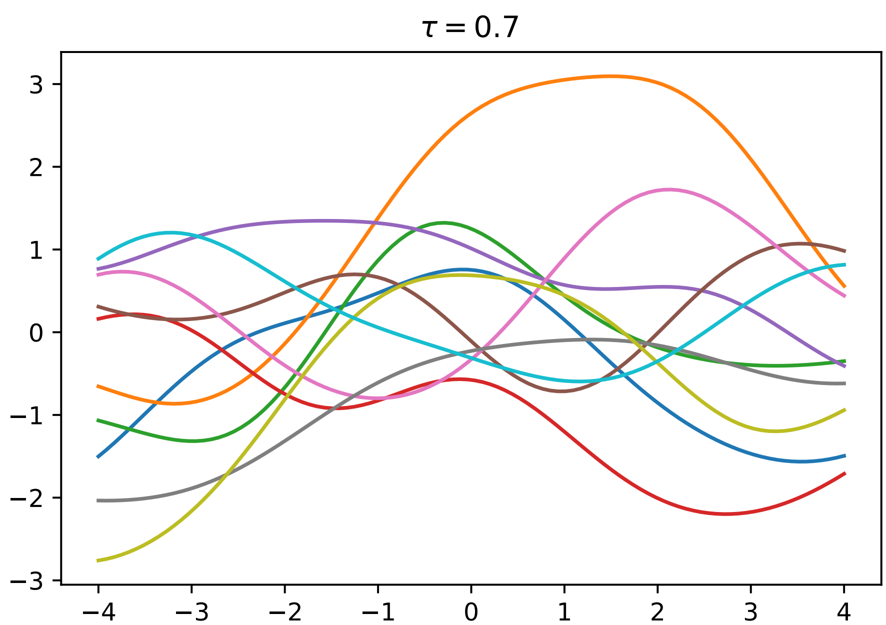
(b) functions become smoother as τ increases



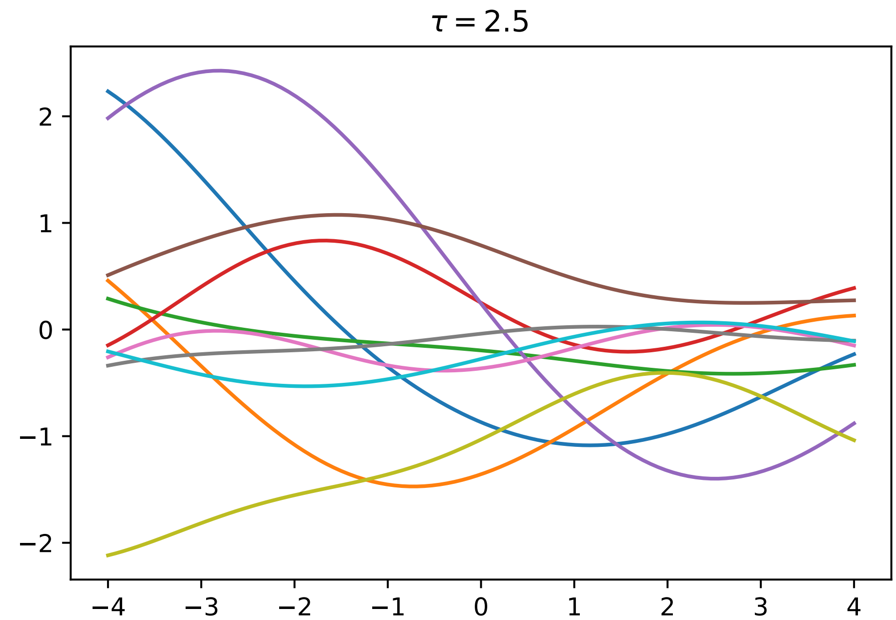




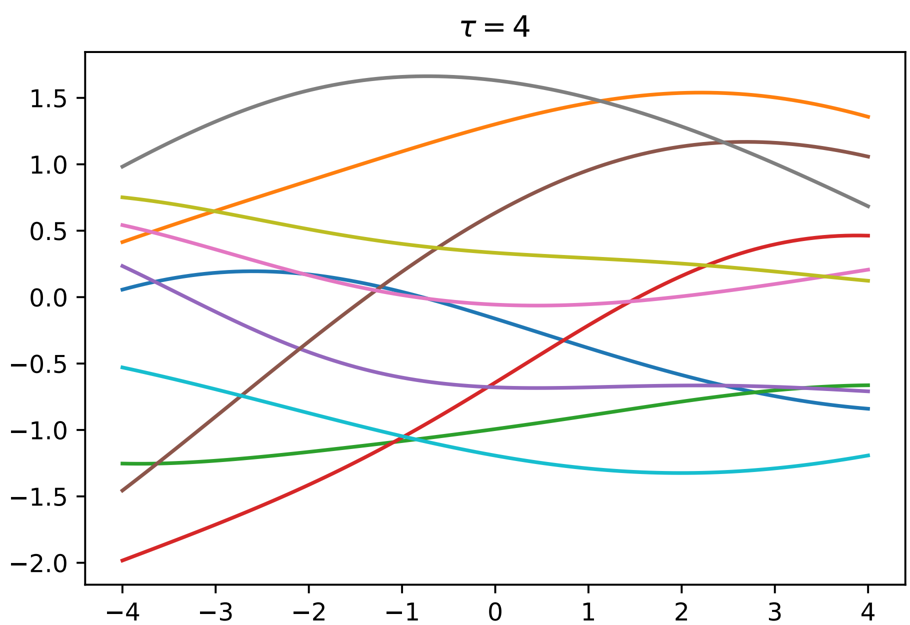
τ = 0.7



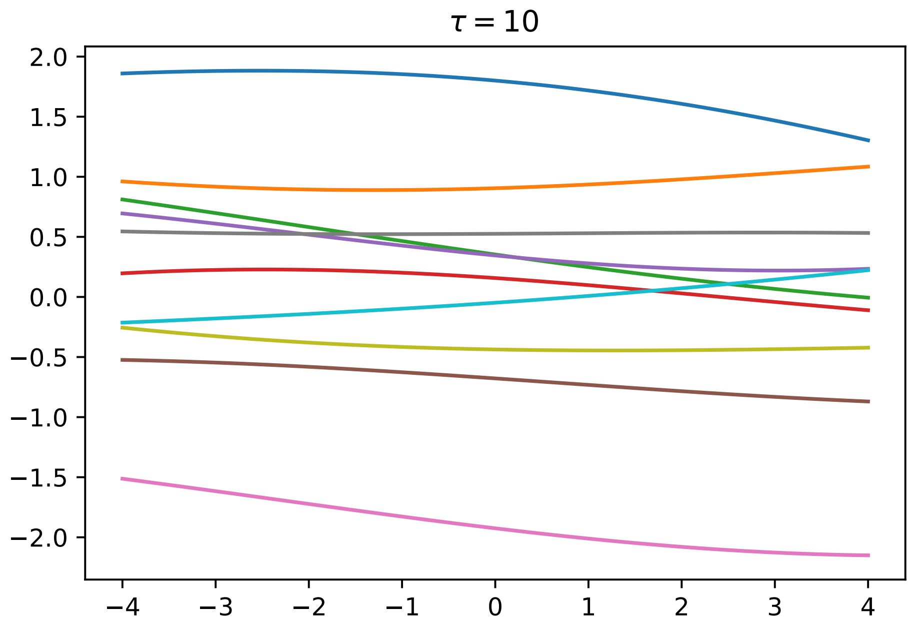
τ = 0.7



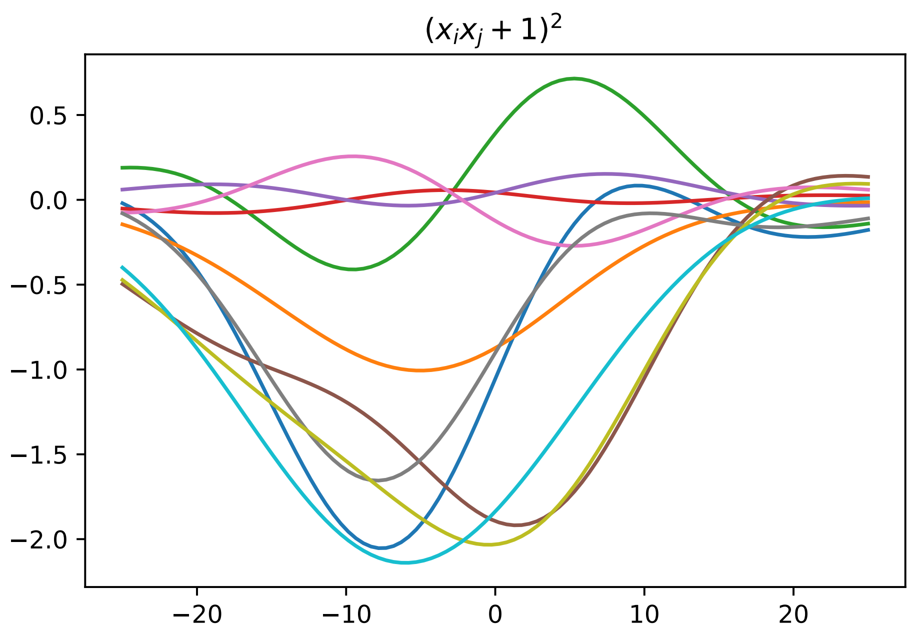
$\tau = 2.5$
2
1
0
−1
−2
−4
−3
−2
−1
0
1
2
3
4



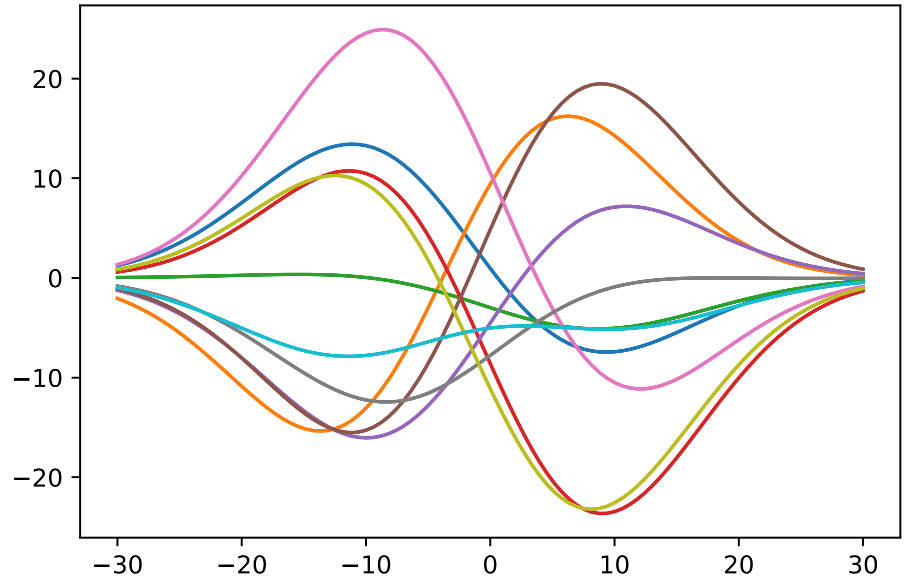
$\tau = 4$
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−4
−3
−2
−1
0
1
2
3
4


$\tau = 10$
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−4
−3
−2
−1
0
1
2
3
4

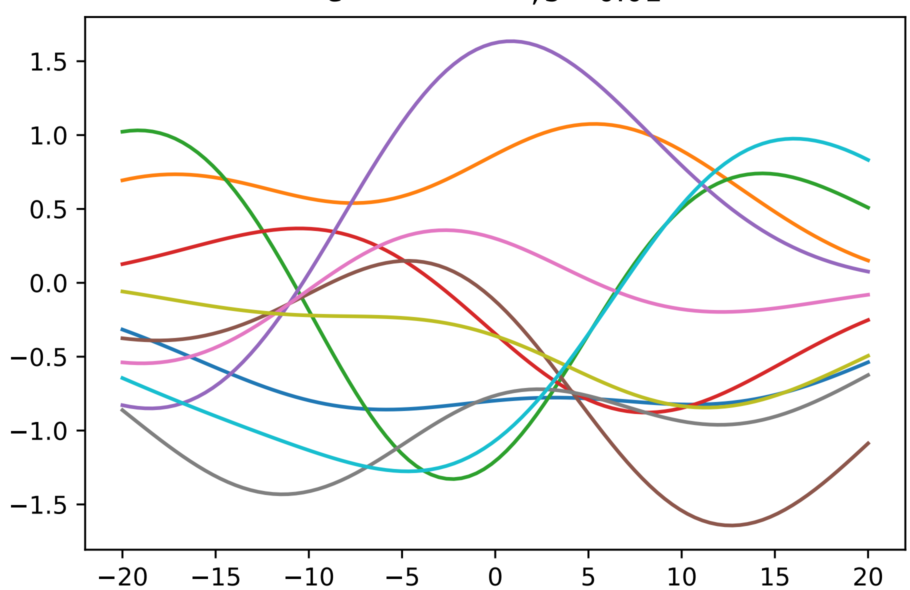
(c)

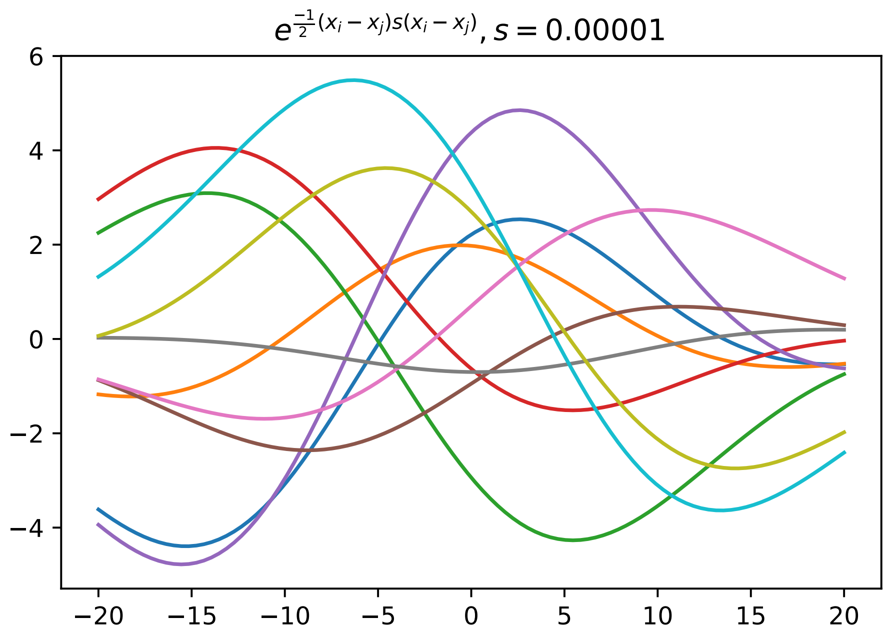

$(x_i x_j + 1)^2$
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−20
−10
0
10
20

$tanh(x_i x_j + 1)$



$e^{\frac{-1}{2}(x_i - x_j)s(x_i - x_j)}, s = 0.01$




$e^{\frac{-1}{2}(x_i-x_j)s(x_i-x_j)}, s = 0.00001$
6
4
2
0
−2
−4
−20
−15
−10
−5
0
5
10
15
20