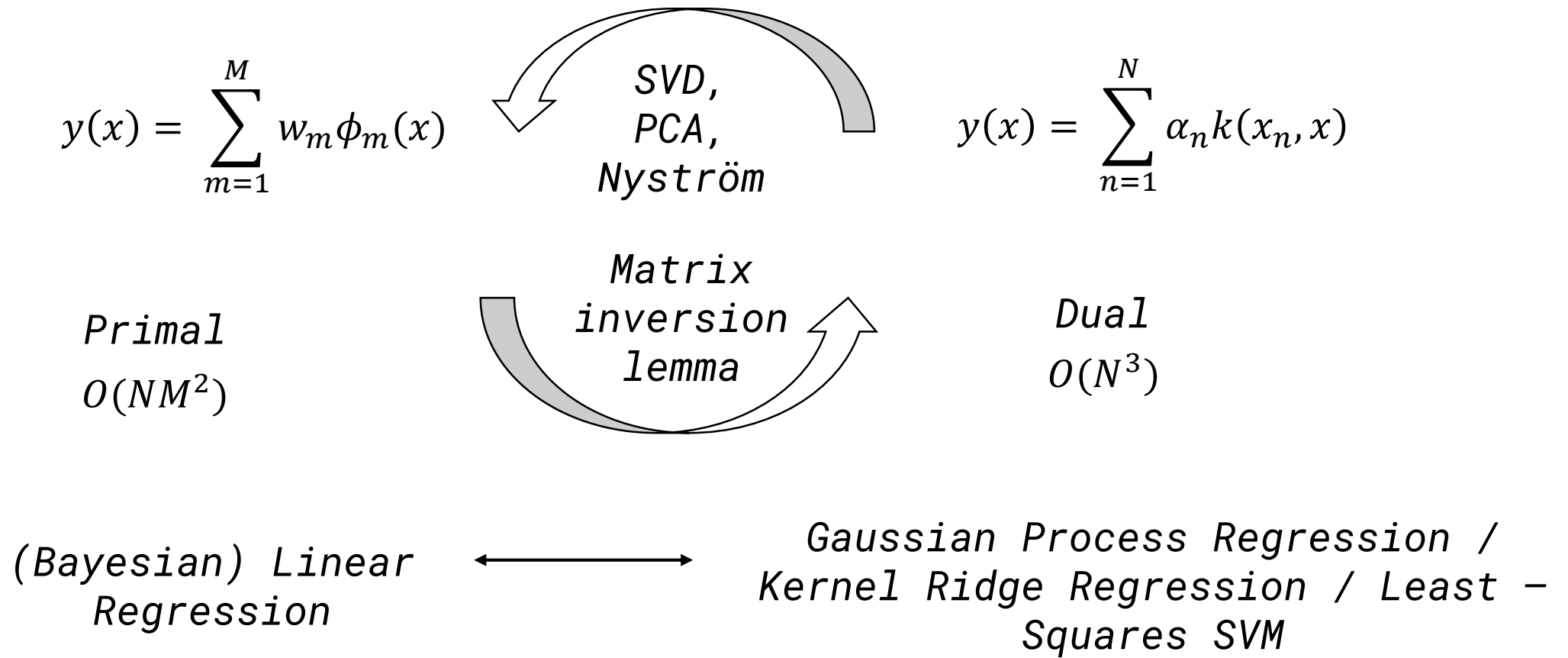


Large-Scale Learning with Fourier Features and Tensor Decompositions

ERNSI Workshop 2021

*Frederiek Wesel, Kim Batselier
Delft University of Technology*

Parametric / Nonparametric linear model

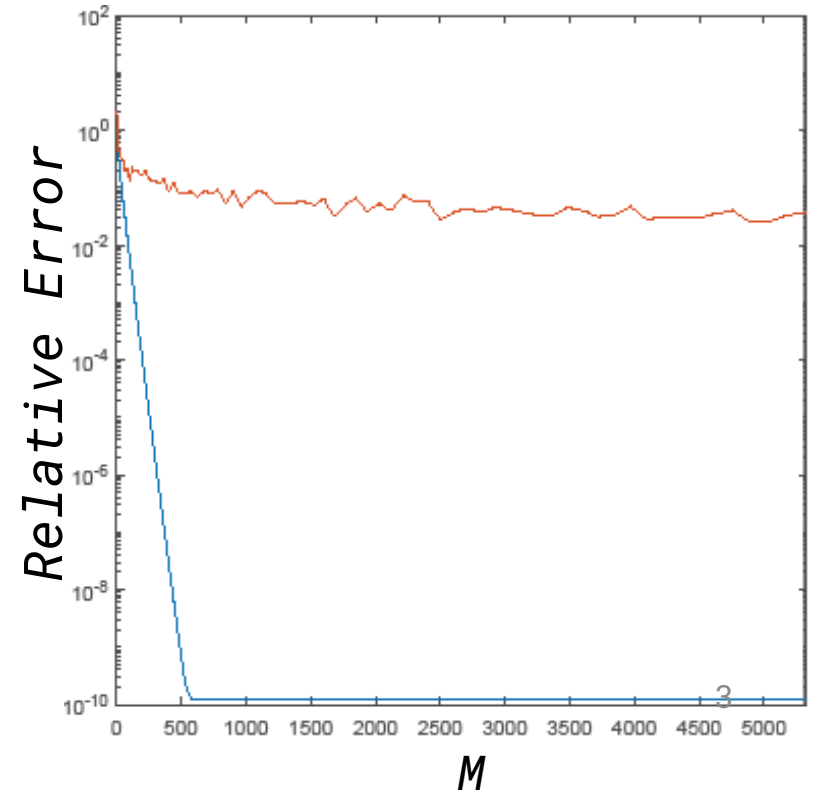


Feature Maps: $k(x, x') \approx \langle \phi(x), \phi(x') \rangle$

$$\phi(x) = \phi_1(x_1) \otimes \phi_2(x_2) \otimes \cdots \otimes \phi_D(x_D)$$

Gaussian Quadrature, Fourier expansions, ... \longrightarrow *Deterministic* Fourier Features, fast *exponential* convergence

Monte Carlo \longrightarrow *Random* Fourier Features, *slow* convergence



Tensor Decompositions

rank- R Canonical Polyadic Decomposition

$$\boldsymbol{\phi}(x) = \phi_1(x_1) \otimes \phi_2(x_2) \otimes \cdots \otimes \phi_D(x_D)$$

\widehat{M}

$$\mathbf{W} = \sum_{r=1}^R w_{1,r} \otimes w_{2,r} \otimes \cdots \otimes w_{D,r}$$

$M = \widehat{M}^D$ terms!

- Unique
- Storage complexity of $O(\widehat{M}RD)$
- R hard to determine

Learning Algorithm

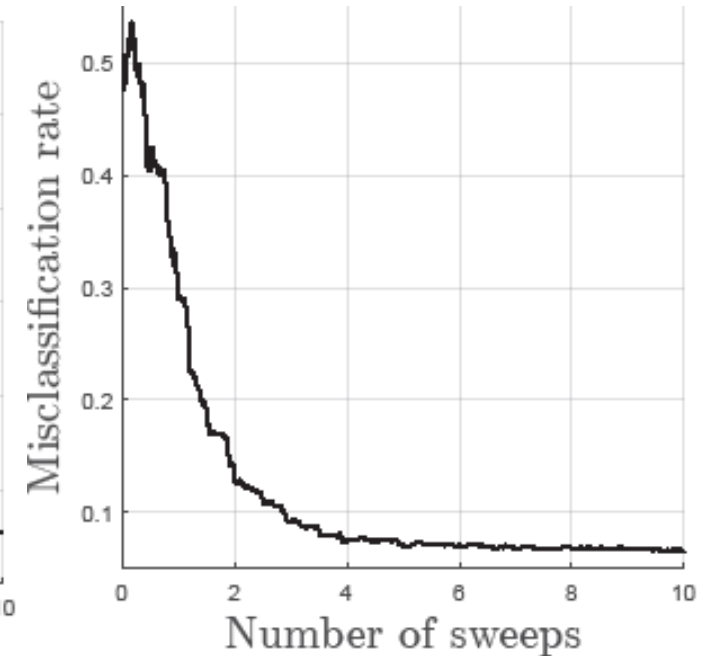
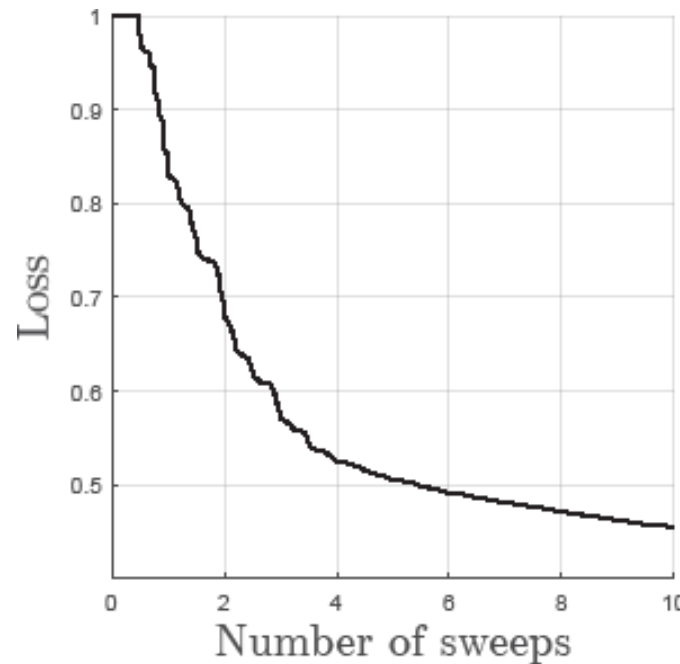
$$\min_{\mathbf{W}} L(\mathbf{W}) = \sum_{n=1}^N (y_n - \langle \mathbf{W}, \boldsymbol{\phi}(x_n) \rangle)^2 + \lambda \langle \mathbf{W}, \mathbf{W} \rangle^2$$

s.t. CP-rank(\mathbf{W}) = R

for $d = 1 \rightarrow D \rightarrow 1 \rightarrow D \rightarrow \dots$

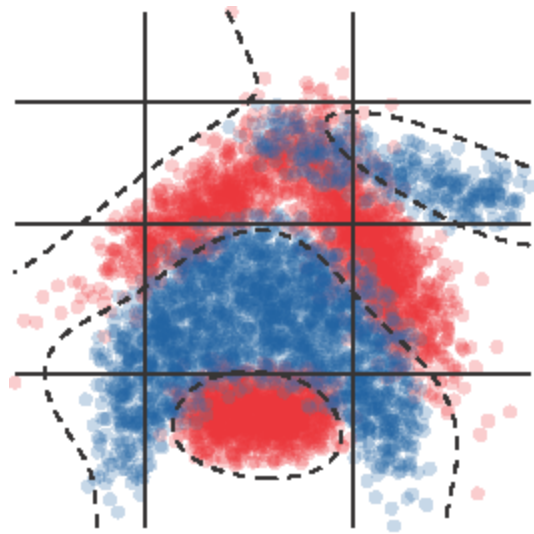
$$\min_{W^{(d)}} L(W^{(d)})$$

- Standard ridge regression
- $O(N\widehat{M}^2R^2)$ flops
- $O(N\widehat{M})$ memory blocks

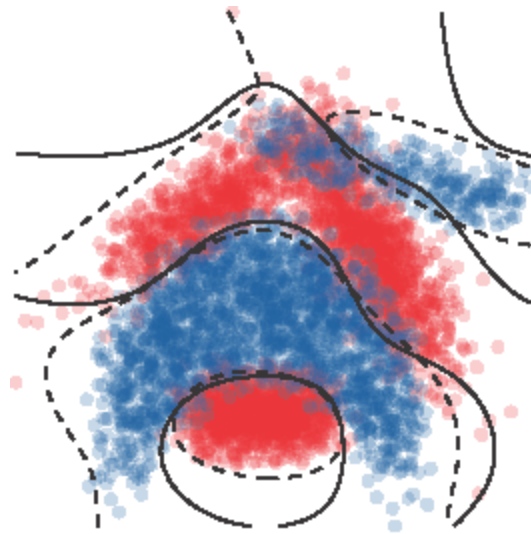


- Classification
- $N = 5300$
- $D = 2$
- $\hat{M} = 12$

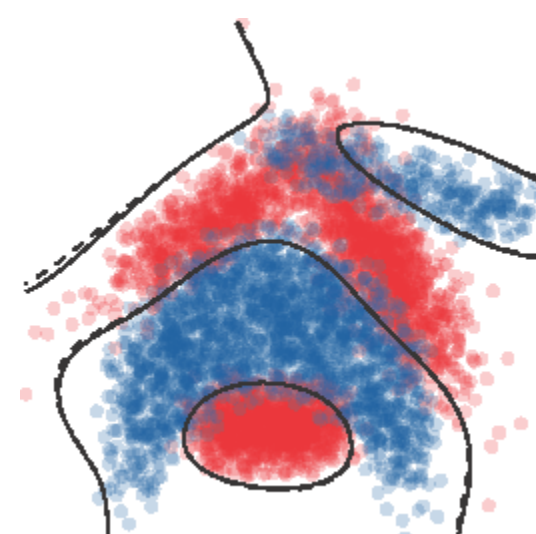
Low-dimensional: Banana Dataset



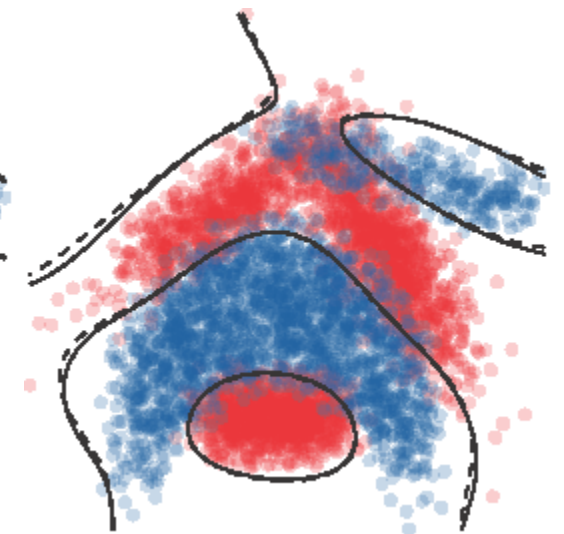
$R = 1$



$R = 2$



$R = 6$



$R = \hat{M} = 12$

- Regression
- $N = 10^4, 10^5, 10^6, \approx 6 \cdot 10^6$
- $D = 8$
- $\hat{M} = 40$
- ≈ 7000 sec on laptop vs ≈ 18600 SVIGP

Large Scale: Airline

N	10000	100000	1000000	5929413
T-KRR($R = 5$)	0.91 ± 0.10	0.82 ± 0.03	0.80 ± 0.02	0.80 ± 0.008
T-KRR($R = 10$)	0.89 ± 0.05	0.80 ± 0.05	0.79 ± 0.02	0.785 ± 0.009
T-KRR($R = 15$)	0.90 ± 0.07	0.80 ± 0.04	0.78 ± 0.02	0.773 ± 0.007
T-KRR($R = 20$)	0.97 ± 0.15	0.78 ± 0.04	0.77 ± 0.01	0.763 ± 0.007
A-Hilbert-GP[Table 1] [38]	0.97 ± 0.14	0.80 ± 0.06	0.83 ± 0.02	0.827 ± 0.005
A-VFF[Table 1] [38]	0.89 ± 0.15	0.82 ± 0.05	0.83 ± 0.01	0.827 ± 0.004
SVIGP[Table 1] [38]	0.89 ± 0.16	0.79 ± 0.05	0.79 ± 0.01	0.791 ± 0.005
VISH[Table 2] [9]	0.90 ± 0.16	0.81 ± 0.05	0.83 ± 0.03	0.834 ± 0.055
GPR [Table 1] [38]	0.89 ± 0.16	N/A	N/A	N/A
A-GPR [Table 1] [38]	0.89 ± 0.16	N/A	N/A	N/A

Thank you for your attention

[https://arxiv.org/abs/2109.
01545](https://arxiv.org/abs/2109.01545)

Frederiek Wesel
f.wesel@tudelft.nl

Kim Batselier
k.batselier@tudelft.nl