Function factorization using warped Gaussian processes

Mikkel N. Schmidt

UNIVERSITY OF CAMBRIDGE

Computational and Biological Learning

Function factorization

- Non-linear regression
  - Input-output points, $\mathcal{D} = \{y^{(n)}, \boldsymbol{x}^{(n)}\}_{n=1}^{N}$
  - Regression function, $y : \mathcal{X} \to \mathbb{R}$
  - Predictions, $p(y^* | \boldsymbol{x}^*, \mathcal{D})$
- Key idea
  - Approximate

    complicated function on high-dimensional space
  - by sum of products of

    simpler functions on subspaces

Motivation

- Function factorization generalizes / combines
  - Matrix and tensor factorization
    Generalized multilinear model
  - Bayesian non-parametric regression
    Warped Gaussian process

Generalized multilinear model

- Describes data as factors
  - Add and multiply any combination of inputs

$$y_{i,j} = f_i^{(1,1)} f_j^{(1,2)} + f_i^{(2,1)} f_j^{(2,2)} + \cdots$$

- Flexible and interpretable

Function factorization model

Components

Factors

$$y(\boldsymbol{x}) = \sum_{k=1}^{K} \prod_{i=1}^{I} f^{(i,k)}(\boldsymbol{x}_i)$$

Output
$y \in \mathbb{R}$

Input
$\boldsymbol{x} \in \mathcal{X}$

Set of functions
$f^{(i,k)} : \mathcal{X}_i \rightarrow \mathbb{R}$

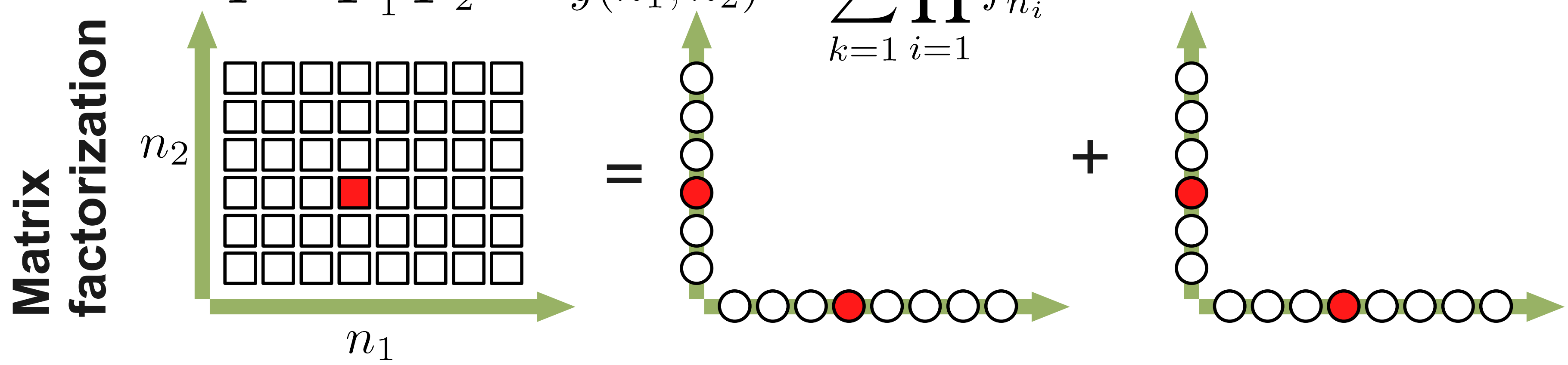
Input in subspace
$\boldsymbol{x}_i \in \mathcal{X}_i \subseteq \mathcal{X}$

Comparison to

Matrix factorization

$$\boldsymbol{Y} = \boldsymbol{F}_1^\top \boldsymbol{F}_2$$

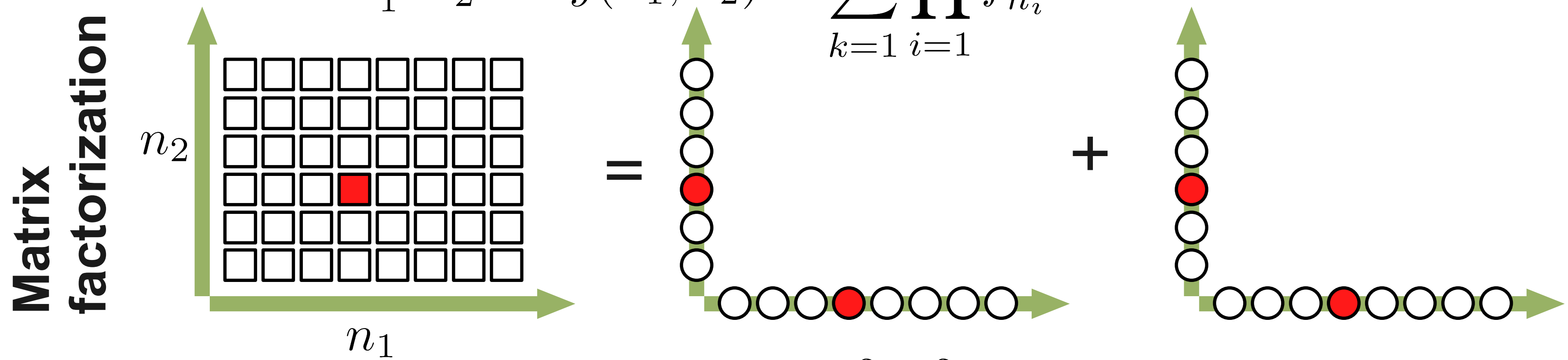
$$y(n_1, n_2) = \sum_{k=1}^{2} \prod_{i=1}^{2} f_{n_i}^{(i,k)}$$



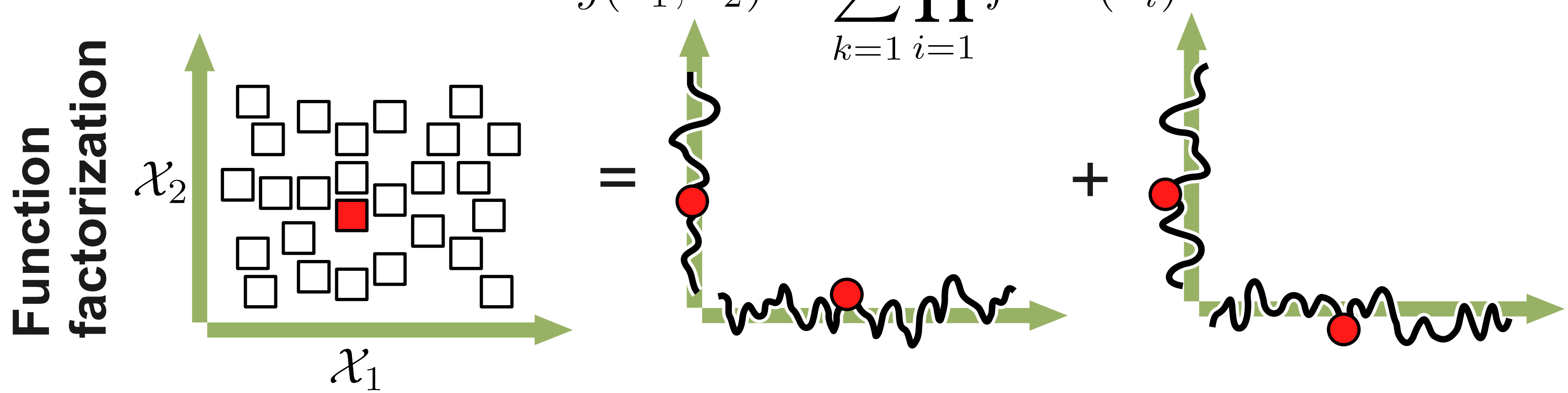
Comparison to

Matrix factorization

$$\boldsymbol{Y} = \boldsymbol{F}_1^\top \boldsymbol{F}_2 \qquad y(n_1, n_2) = \sum_{k=1}^{2} \prod_{i=1}^{2} f_{n_i}^{(i,k)}$$

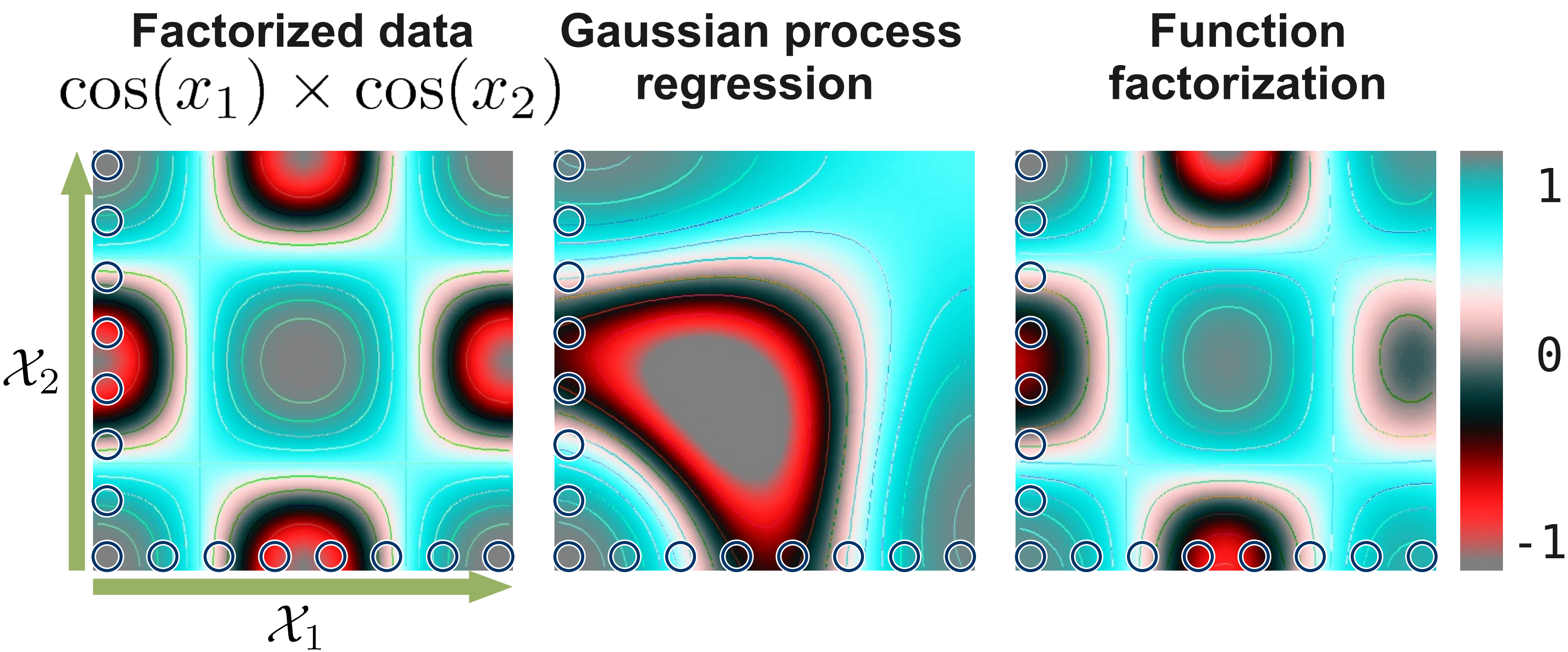


$$y(x_1, x_2) = \sum_{k=1}^{2} \prod_{i=1}^{2} f^{(i,k)}(x_i)$$



Comparison to

Gaussian process regression



Priors over functions

1) Parametric functions
   - Limited flexibility
2) Gaussian processes
   - Flexible and non-parametric
   - Limited by joint Gaussianity assumption
3) Warped Gaussian processes
   - GP warped by non-linear function

Warped Gaussian processes Snelson et al. (1999)

- GP warped by non-linear function

$$f(\boldsymbol{x}) = h(g(\boldsymbol{x})) \qquad g(\boldsymbol{x}) \sim \mathcal{GP}\big(m(\boldsymbol{x}), c(\boldsymbol{x}, \boldsymbol{x}')\big)$$

Non-linear warp function

Gaussian process

Mean function

Covariance function

- Parameters of warp and covariance functions are learned from data

Inference

- Hamiltonian Markov chain Monte Carlo
  Duane et al. (1987)
- Integrate out all parameters
  - Likelihood function (noise variance)
  - GP latent variables
  - Covariance functions
  - Warp functions
- Gradients wrt. all parameters

Color of beef data

Bro and Jakobsen (2002)

Color of beef as it changes during storage

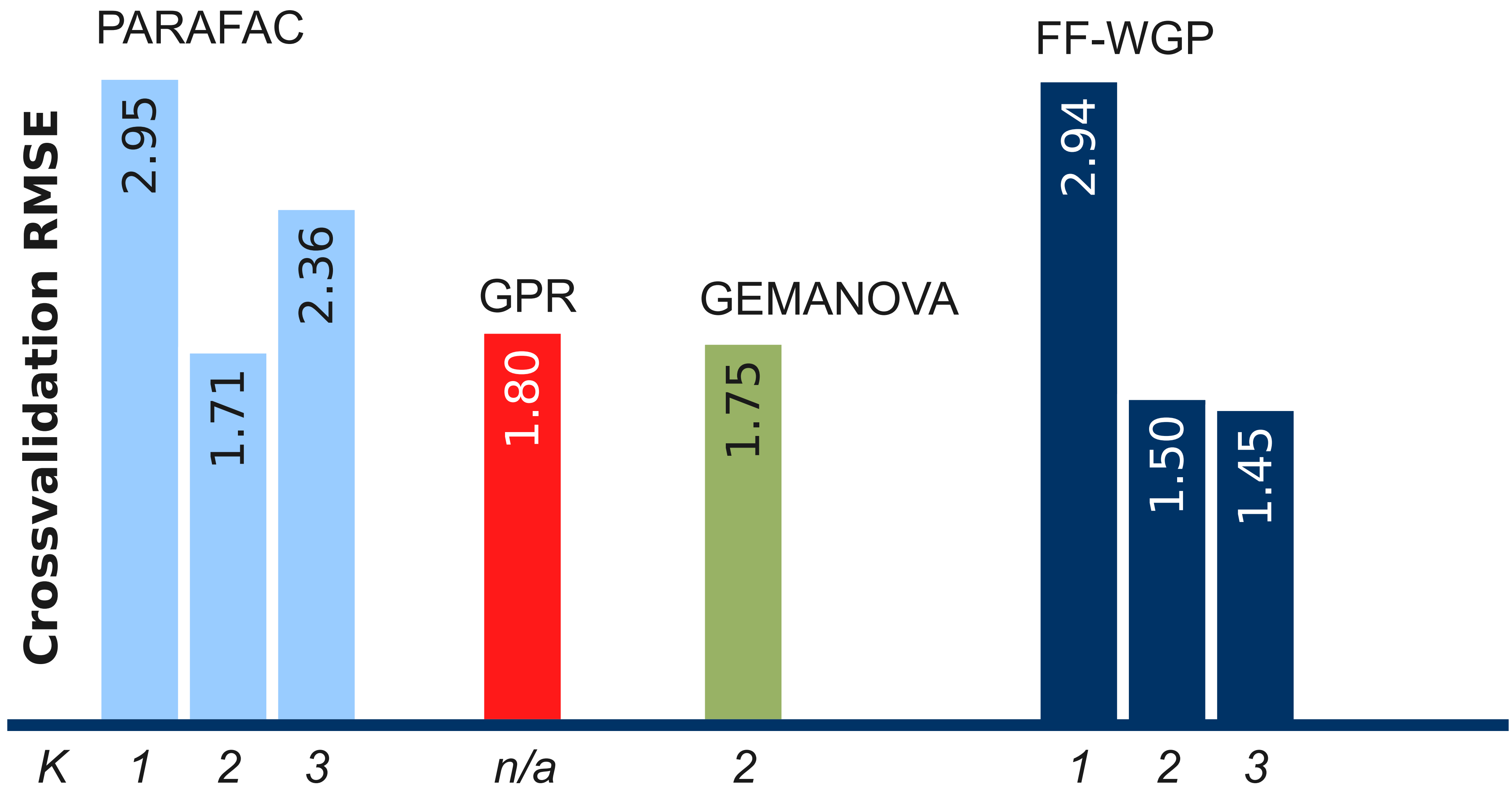
- Storage time
- Temperature
- Oxygen content
- Exposure to light

Task: Predict color from condition

Color of beef data

- Data: 5-way array
  - Measured red color on non-negative scale
- 60% missing val
  - PARAFAC: Handle missing data using EM iterations
  - Function factorization: Does not require data on grid
- Warp function
  - Parameterized function that maps to the non-negative numbers

Results



PARAFAC
GPR
GEMANOVA
FF-WGP
Crossvalidation RMSE
2.95
1.71
2.36
1.80
1.75
2.94
1.50
1.45
K
1
2
3
n/a
2
1
2
3

Summary

- New approach to non-linear regression
- Generalizes matrix and tensor factorization
- Exploits factorized structure in data
- Warped Gaussian process priors over functions
- Bayesian inference (Hamiltonian Monte Carlo)
  - Integrate out all parameters
- Outperforms PARAFAC and GPR

References

Bro and Jakobsen (2002), Exploring complex interactions in designed data using GEMANOVA. Color changes in fresh beef during storage. Chemometrics, Journal of, 16, 294–304.

Duane, et al. (1987), Hybrid Monte Carlo. Physics Letters B, 195, 216–222.

Rasmussen and Williams (2006), Gaussian processes for machine learning. MIT Press.

Snelson et al. (2004), Warped gaussian processes. Neural Information Processing Systems, Advances in (NIPS) (pp. 337–344).