Hello! We are the team “Flash”

Sascha Gaudlitz
HU Berlin

Konstantins Starovoitovs
HU Berlin

Mike Nguyen
TU Braunschweig

Tim Sieker
CAU Kiel

Eric Ziebell
HU Berlin

Part 1: Normalising Flows – Theory

Change of variable formula

$$\mathbf{z} \sim \pi(\mathbf{z}), \mathbf{x} \sim p(\mathbf{x}), \mathbf{x} := f(\mathbf{z}), \mathbf{z} = f^{-1}(\mathbf{x})$$

$$p(\mathbf{x}) = \pi(\mathbf{z}) \left| \det \frac{d\mathbf{z}}{d\mathbf{x}} \right| = \pi\left(f^{-1}(\mathbf{x})\right) \left| \det \frac{df^{-1}}{d\mathbf{x}} \right|$$

Normalizing flows

composition of easily invertible bijective transformations to map a simple distribution to a more complex one

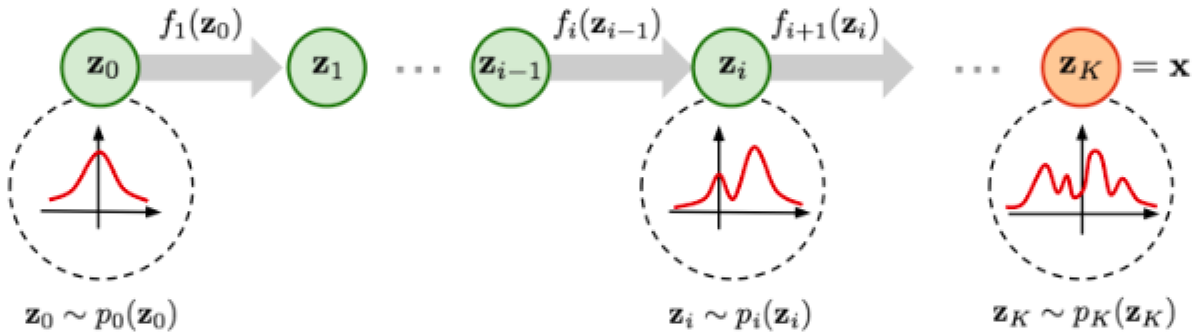
$$\mathbf{z}_{i-1} \sim p_{i-1}(\mathbf{z}_{i-1}), \mathbf{z}_i := f_i(\mathbf{z}_{i-1}), \mathbf{z}_{i-1} = f_i^{-1}(\mathbf{z}_i)$$

$$p_i(\mathbf{z}_i) = p_{i-1}\left(f_i^{-1}(\mathbf{z}_i)\right) \left| \det \frac{df_i^{-1}}{d\mathbf{z}_i} \right| = p_{i-1}(\mathbf{z}_{i-1}) \left| \det \frac{df_i}{d\mathbf{z}_{i-1}} \right|^{-1}$$

inverse fct. thm.
determinant of the inverse

$$\mathbf{x} := \mathbf{z}_K = f_K \circ f_{K-1} \circ \cdots \circ f_1(\mathbf{z}_0)$$

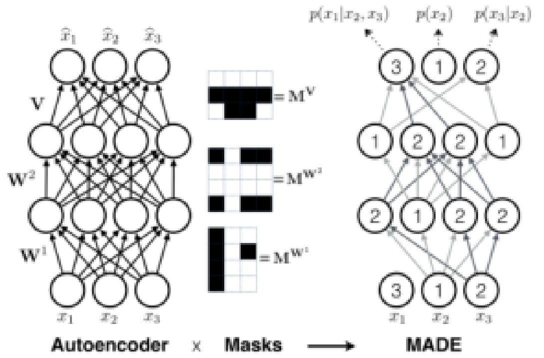
- directly modelling pdf the of real data ⇨ tractable likelihood
- generating data with simple transformations
- maximize for log-likelihood



Autoregressive models

$$p(\mathbf{x}) = \prod_{i=1}^{D} p(x_i | x_1, \ldots, x_{i-1}) = \prod_{i=1}^{D} p(x_i | x_{1:i-1})$$

Masked Autoencoder for Density Estimation (MADE)



- autoregressive architecture for the above MAF
- allows computation of the autoregressive location and scale parameters in one forward pass, by application of binary masks on weight matrices

Masked autoregressive flow (MAF)

Data generation:

$$x_i \sim p(x_i | \mathbf{x}_{1:i-1}) = z_i \odot \sigma_i(\mathbf{x}_{1:i-1}) + \mu_i(\mathbf{x}_{1:i-1}), \text{ where } \mathbf{z} \sim \pi(\mathbf{z})$$

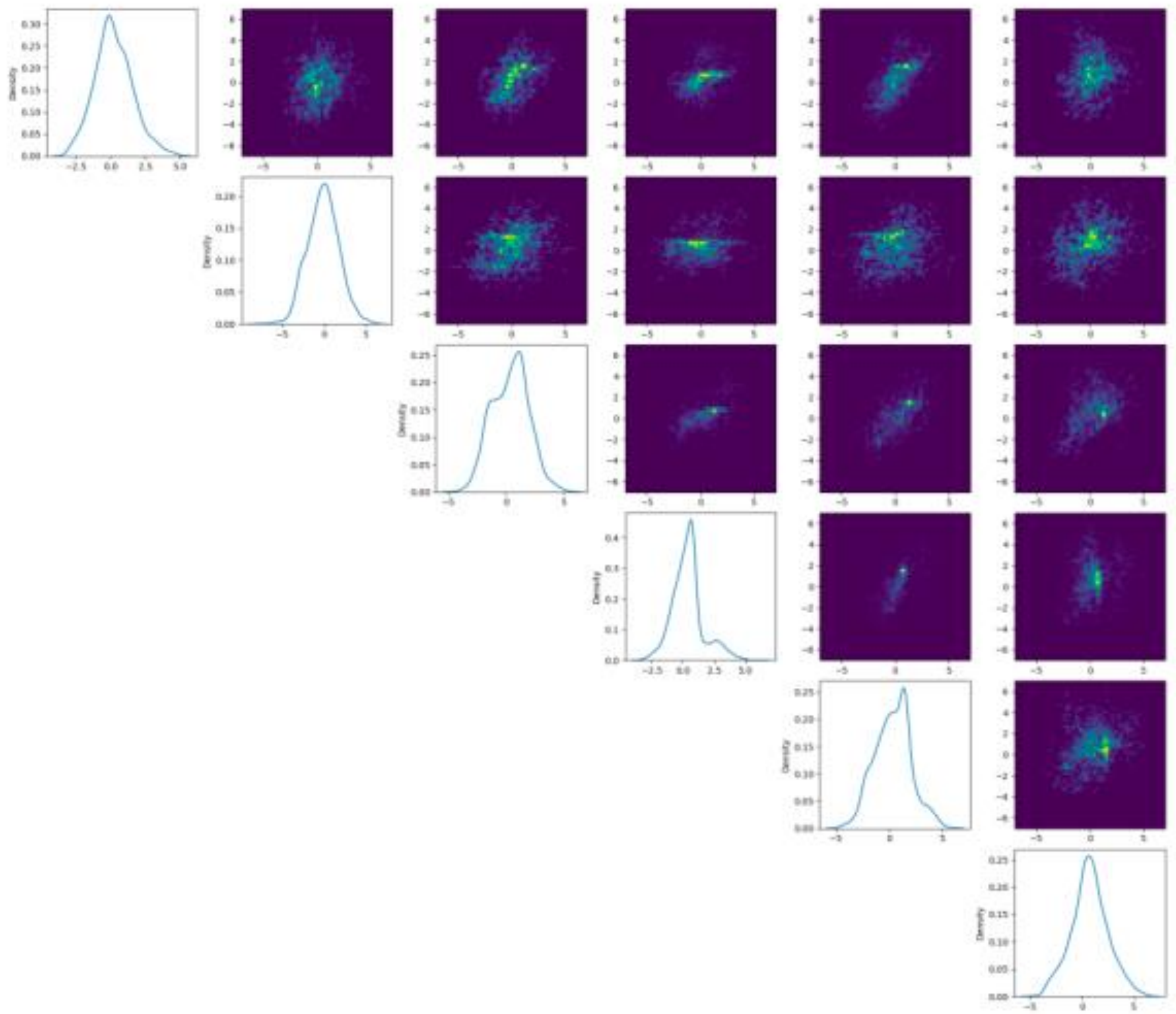
Density estimation:

$$p(\mathbf{x}) = \prod_{i=1}^{D} p(x_i | \mathbf{x}_{1:i-1})$$

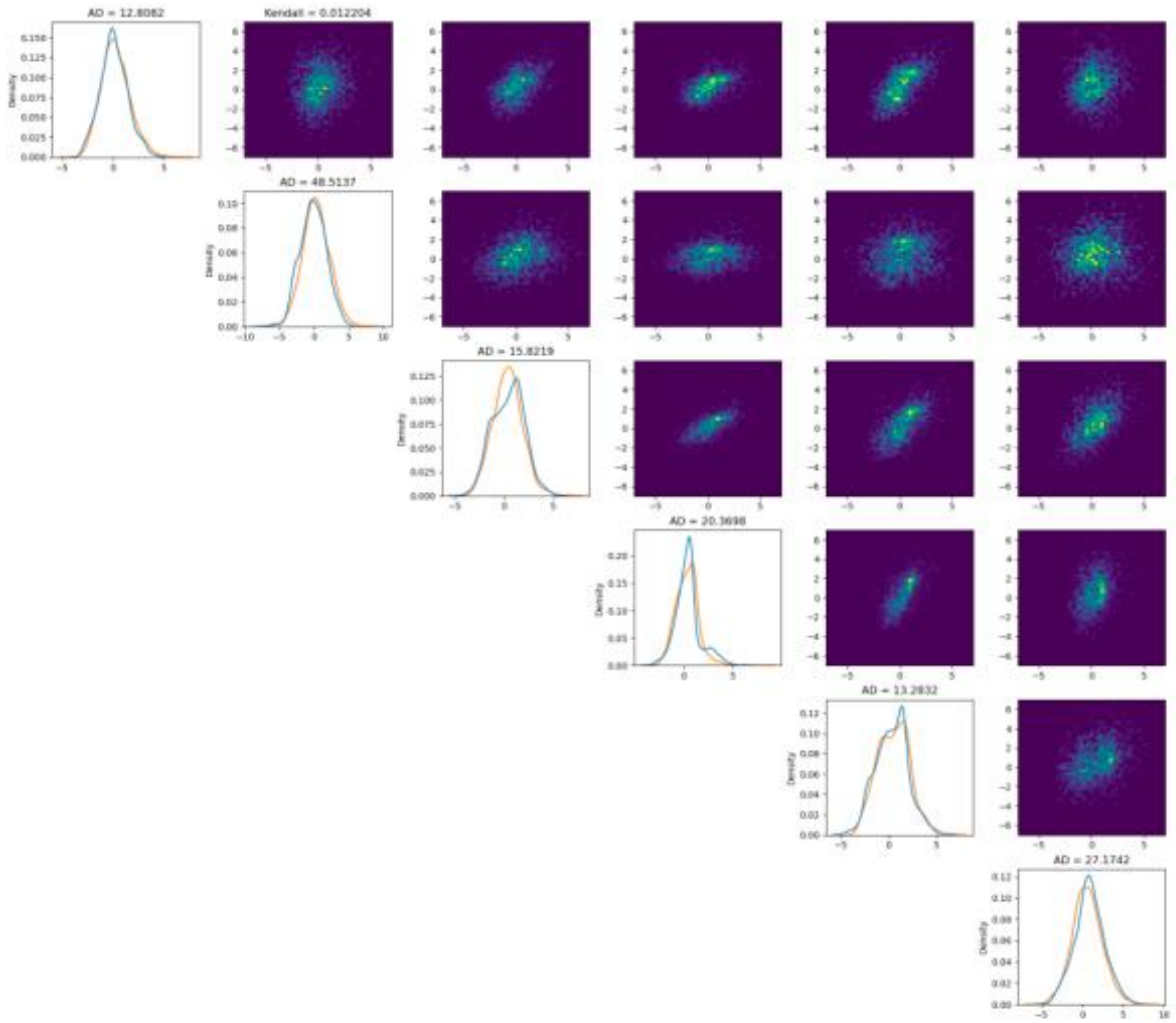
- affine transformation ⇨ elementary inverse and Jacobian determinant
- location and scale given in terms of neural nets
- backprop with the loss given by negative log-likelihood
- ad-hoc incorporation linear trend and learnable sample weights

- Germain, Mathieu, et al. "Made: Masked autoencoder for distribution estimation." International conference on machine learning. PMLR, 2015.
- Papamakarios, George, Theo Pavlakou, and Iain Murray. "Masked autoregressive flow for density estimation." (2017).
- Weng, L. (2018). Flow-based Deep Generative Models. lilianweng.github.io.

Part 1: Normalising Flows – Results



Marginals of the real data



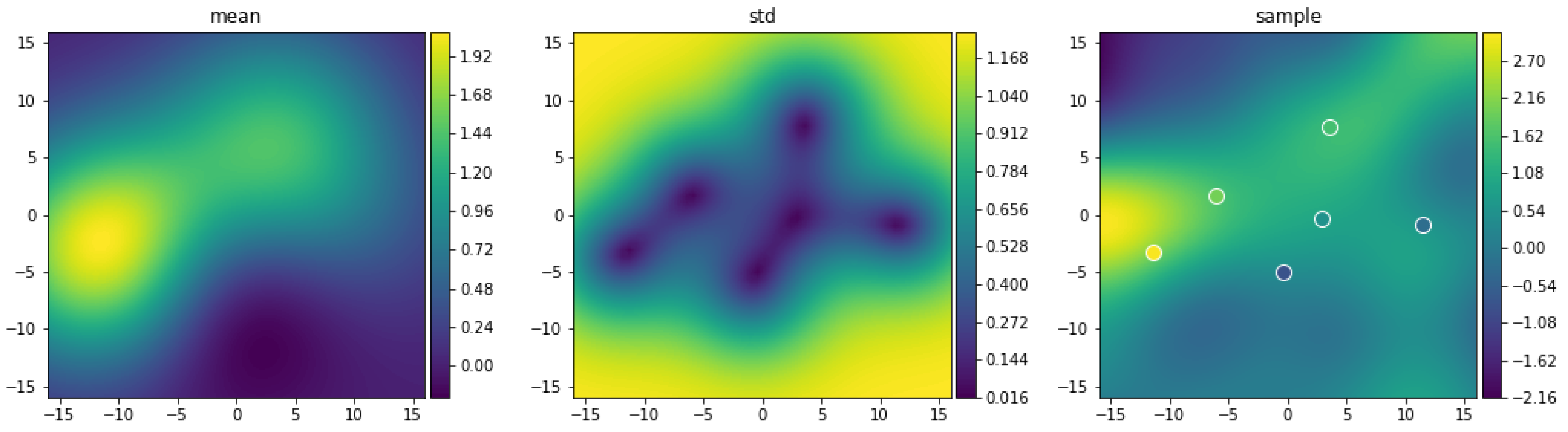
Marginals of the generated data

Part 2: Gaussian Process Regression (Kriging)

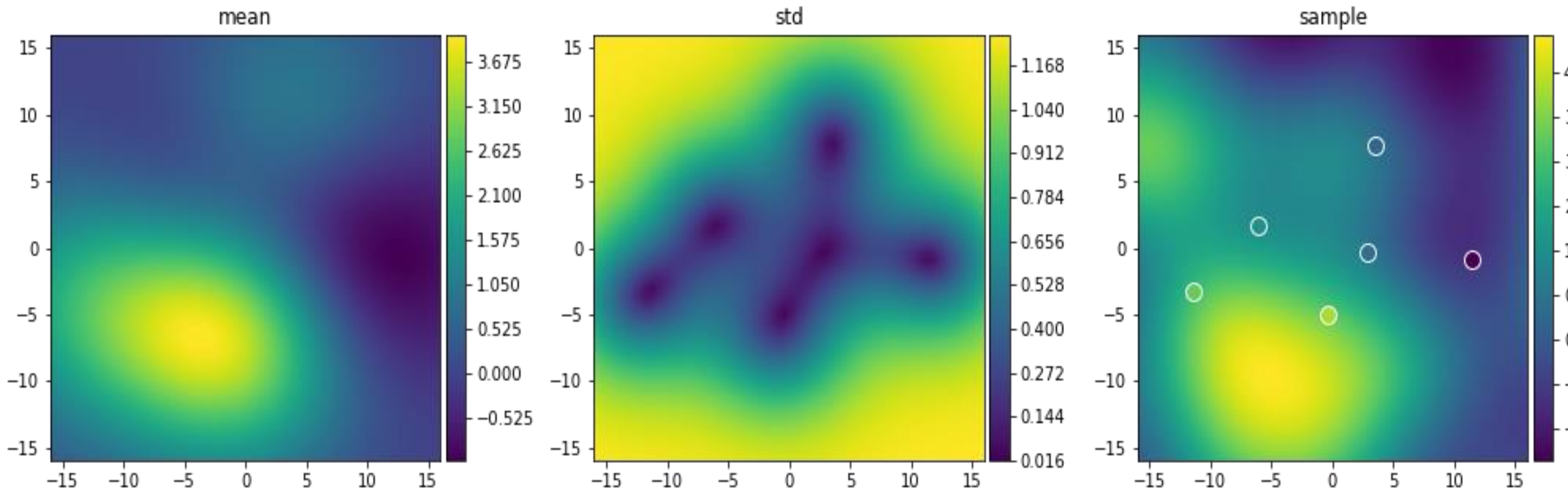
For admissible kernel $k\colon \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ set

$$K(U,V) = (k(u_i, v_j))_{\substack{i=1,\dots,n_u \\ j=1,\dots,n_v}}$$

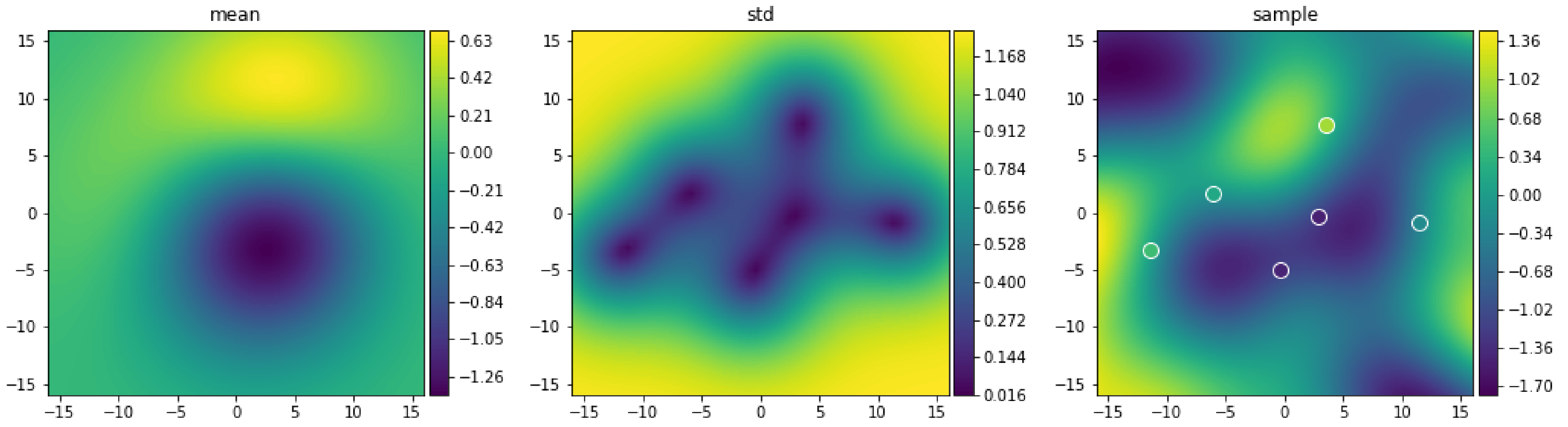
$$\mathbf{f}_* | X_*, X, \mathbf{f} \;\sim\; \mathcal{N}\big(K(X_*,X)K(X,X)^{-1}\mathbf{f},\\ K(X_*,X_*) - K(X_*,X)K(X,X)^{-1}K(X,X_*)\big)$$



01-01-2012



01-01-2016



01-01-2008

C. E. Rasmussen & C. K. I. Williams, Gaussian Processes for Machine Learning, the MIT Press, 2006, ISBN 026218253X. © 2006 Massachusetts Institute of Technology. www.GaussianProcess.org/gpml