Gaussian Process Latent Variable Models

Rajbir-Singh Nirwan

22.01.2018

FIAS Frankfurt Institute for Advanced Studies

Overview

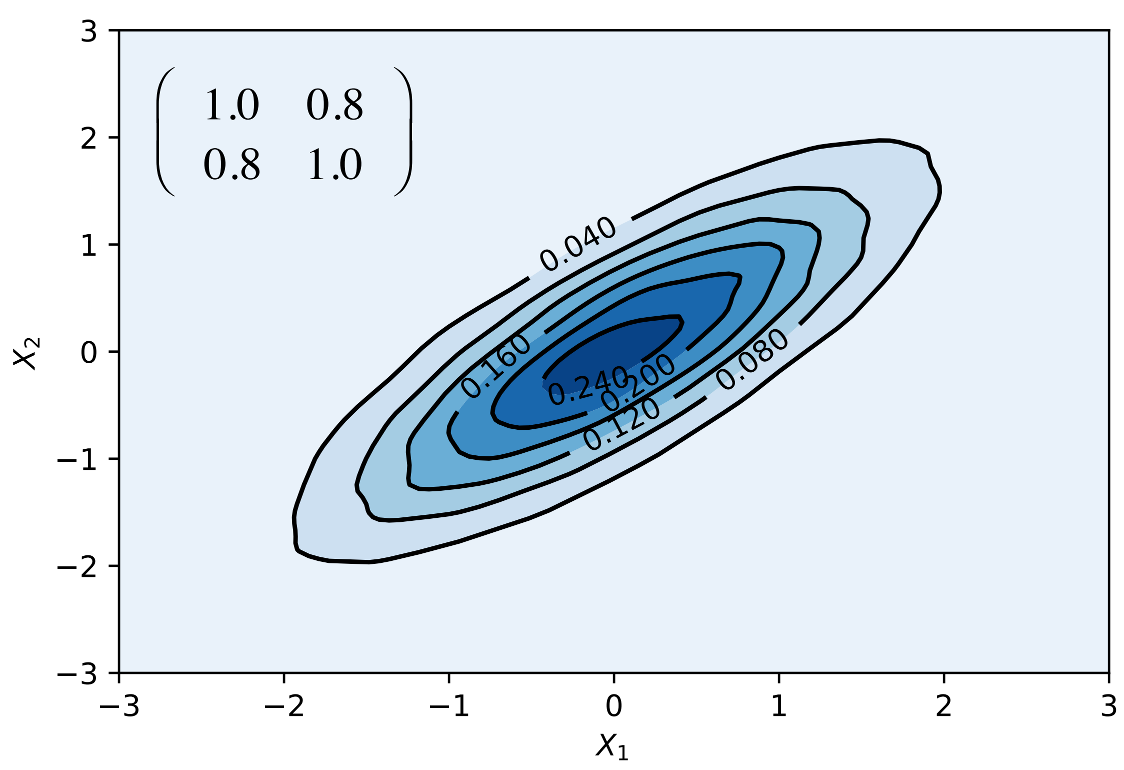
- Gaussian Processes (GPs)
- Latent Variable Models (LVMs)
- Gaussian Process Latent Variable Models (GP-LVMs)
- Applications

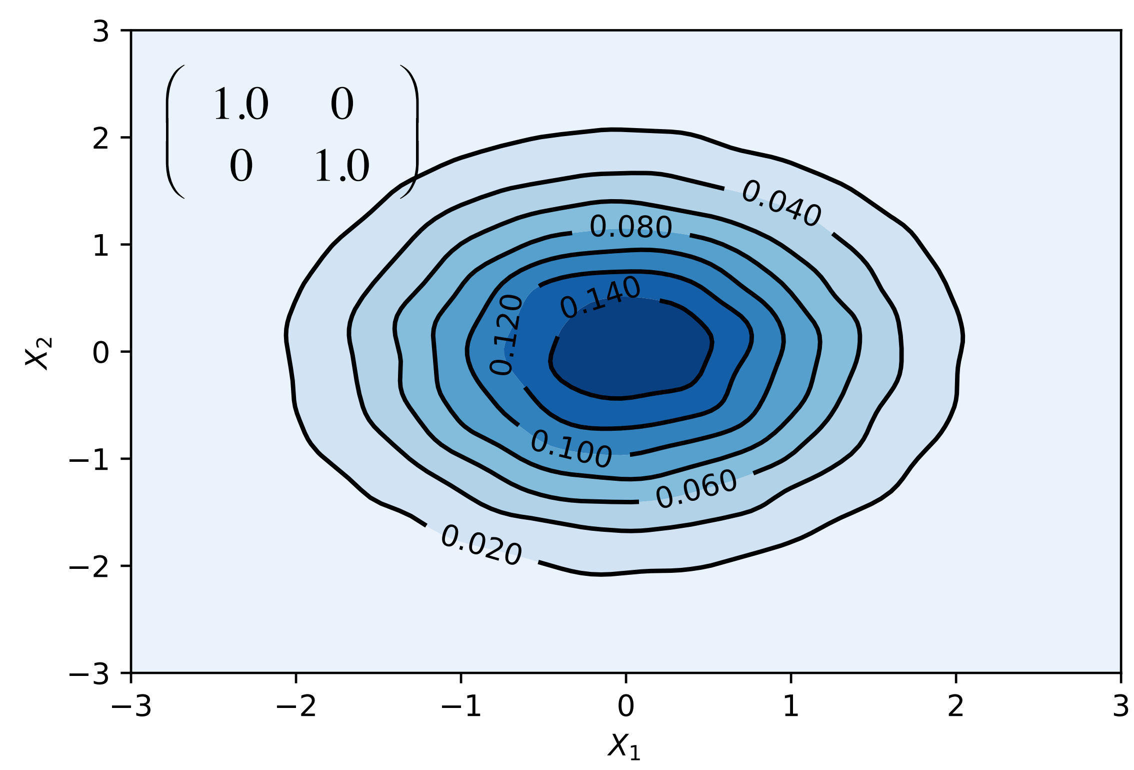
Gaussian Distribution

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Multivariate Gaussian Distribution

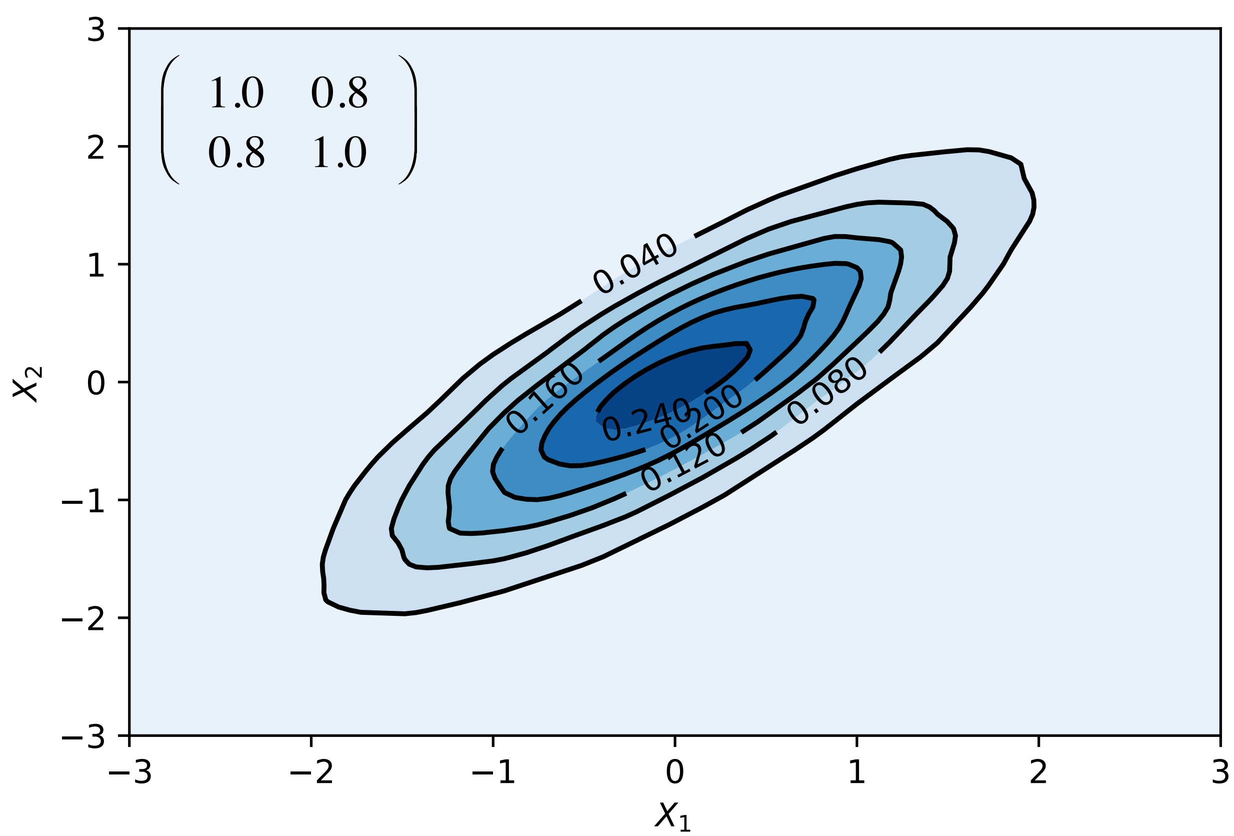
$$p(x_1,\ldots,x_D) = \frac{1}{\sqrt{(2\pi)^D\,|\Sigma|}}\, e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$$

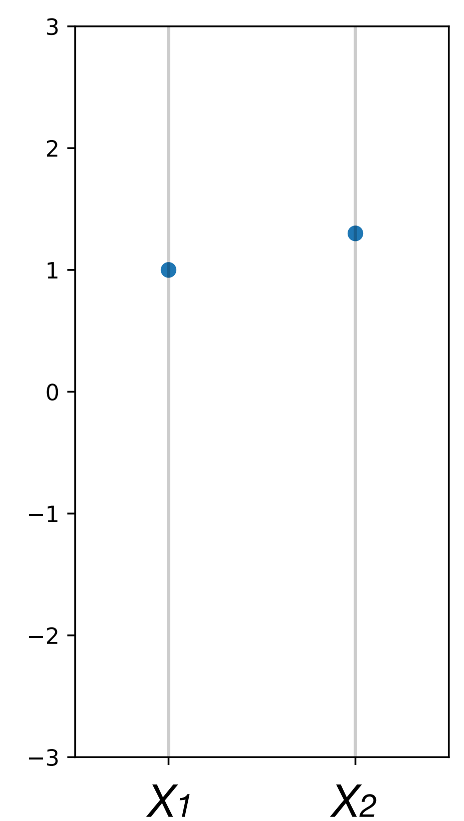




Multivariate Gaussian Distribution

$$p(x_1,\ldots,x_D) = \frac{1}{\sqrt{(2\pi)^D\,|\Sigma|}}\, e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)}$$

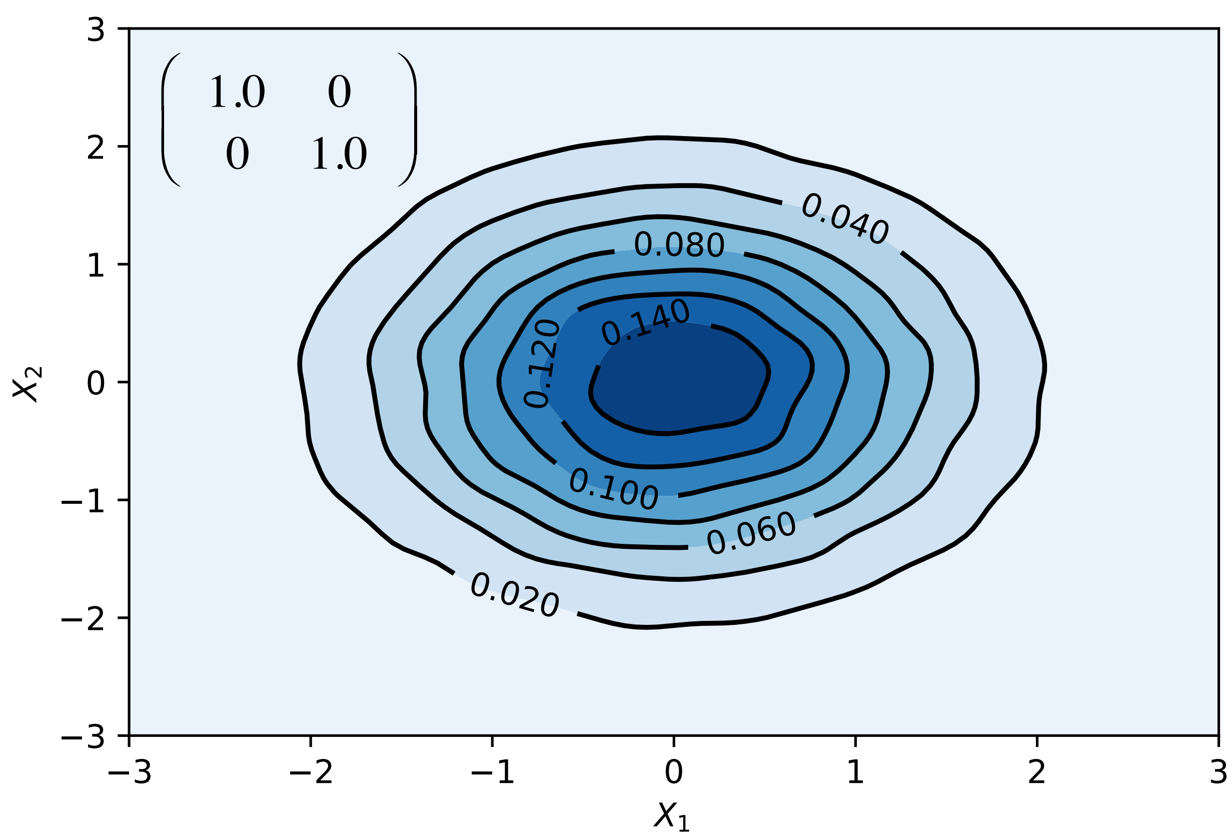


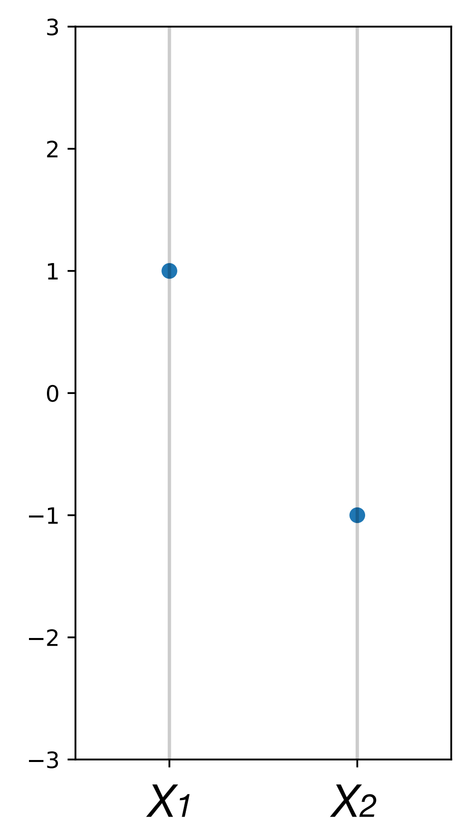


$$k(x_1, x_2) = 0.8$$

Multivariate Gaussian Distribution

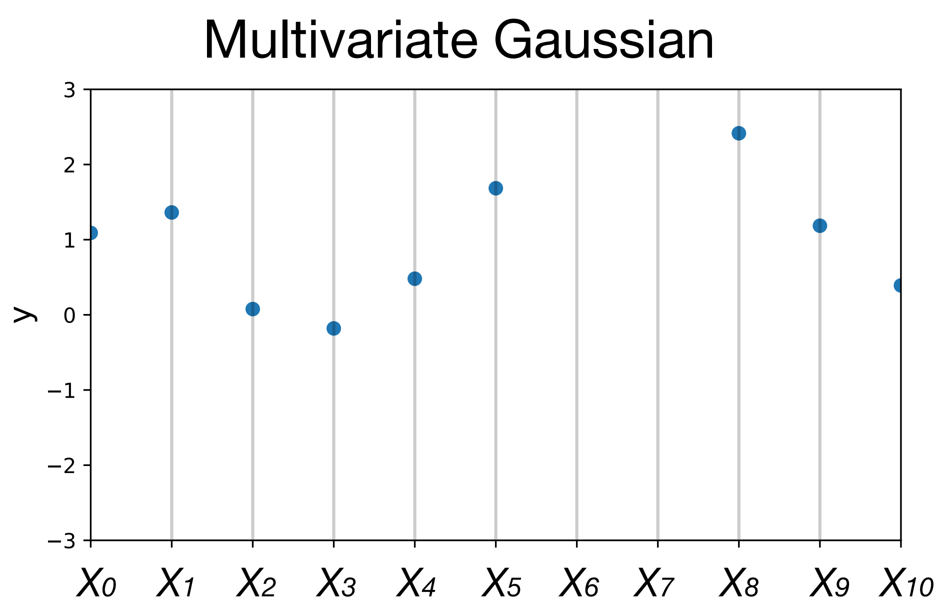
$$p(x_1,\ldots,x_D) = \frac{1}{\sqrt{(2\pi)^D |\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1} (x-\mu)}$$





$$k(x_1, x_2) = 0$$

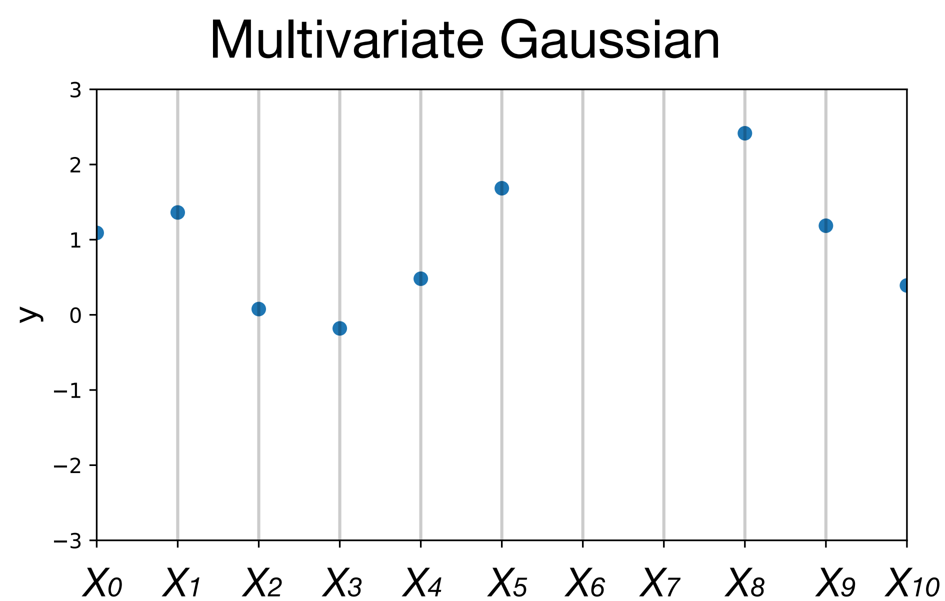
Multivariate Gaussian Distribution



$$Y \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} k_{00} & \cdots & k_{0N} \\ \vdots & \ddots & \vdots \\ k_{N0} & \cdots & k_{NN} \end{pmatrix} \right)$$

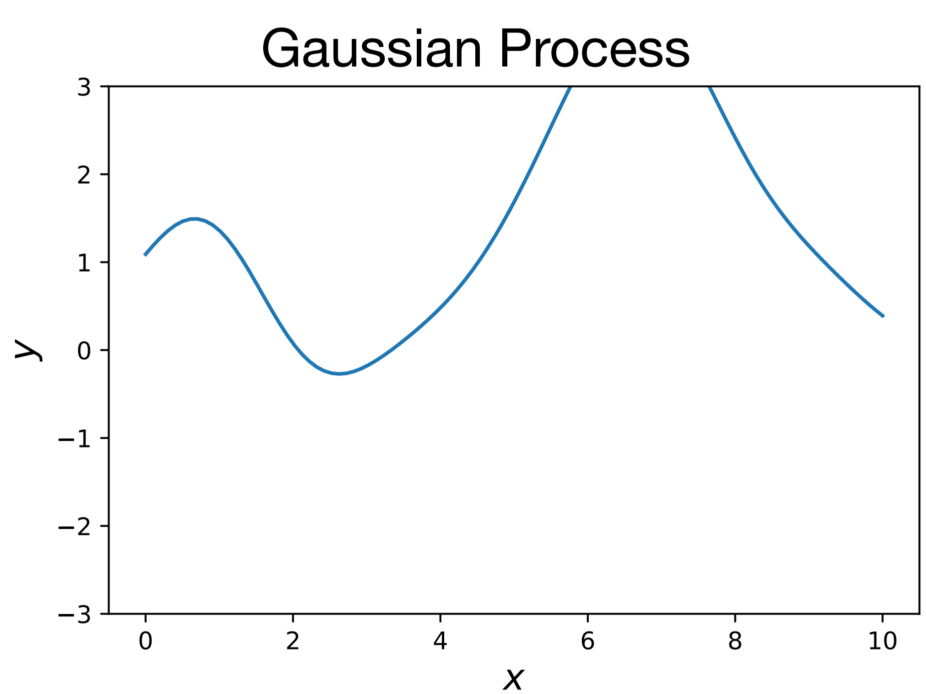
$$k(x_i, x_j) = e^{-\frac{1}{2}(x_i - x_j)^2}$$

Gaussian Processes



$$Y \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} k_{00} & \cdots & k_{0N} \\ \vdots & \ddots & \vdots \\ k_{N0} & \cdots & k_{NN} \end{pmatrix}\right)$$

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



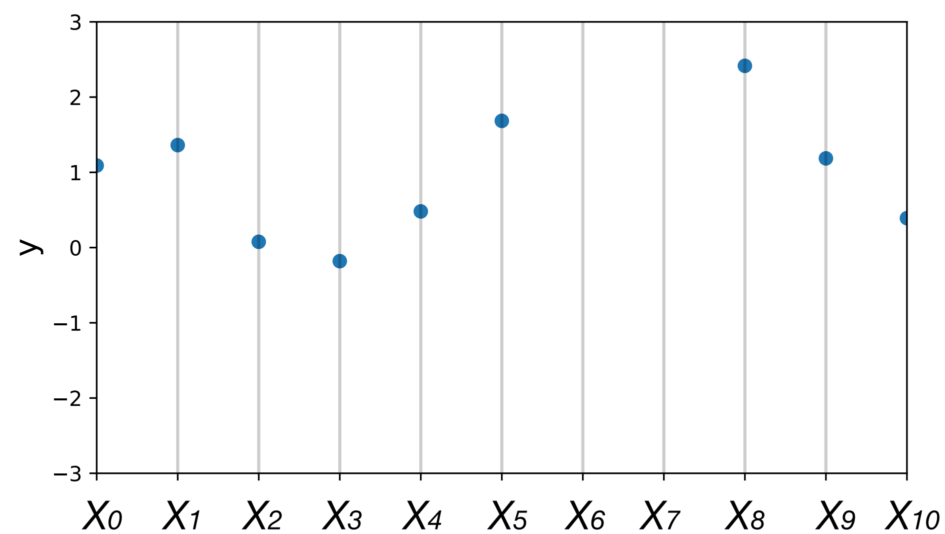
$$Y \sim GP\big(m(x), k(x, x')\big)$$

$$m(x) = 0$$

$$k(x, x') = e^{-\frac{1}{2}(x - x')^2}$$

Gaussian Processes

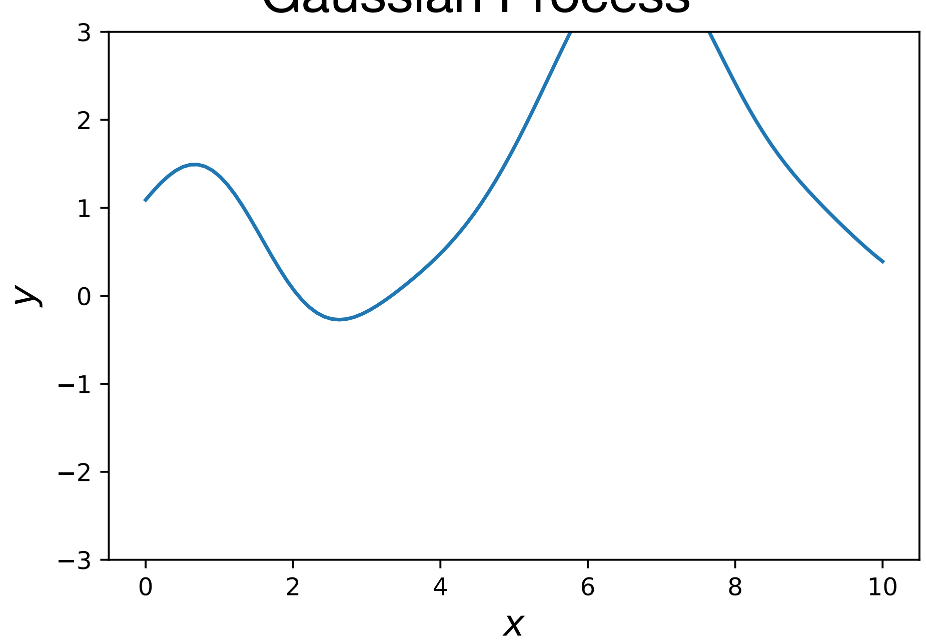
Multivariate Gaussian



$$Y \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} k_{00} & \cdots & k_{0N} \\ \vdots & \ddots & \vdots \\ k_{N0} & \cdots & k_{NN} \end{pmatrix}\right)$$

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

Gaussian Process



Other Kernels

Linear $\quad k_\theta(x_i, x_j) = \theta x_i x_j$

RBF $\quad k_\theta(x_i, x_j) = \theta_1 e^{-\frac{1}{2\theta_2}(x_i - x_j)^2}$

OU $\quad k_\theta(x_i, x_j) = \theta_1 e^{-\frac{1}{2\theta_2}|x_i - x_j|}$

Periodic $\quad k_\theta(x_i, x_j) = \theta_1 e^{-\frac{1}{2\theta_2}\left(sin^2(\theta_3(x_i - x_j))\right)}$

Bayesian Machine Learning

- We are living in a really simple world
- Only known (data) and unknown (hypothesis) quantities exist

$$P(hypothesis \mid data) = \frac{P(data \mid hypothesis)P(hypothesis)}{P(data)}$$

Bayesian Machine Learning

- We are living in a really simple world
- Only known (data) and unknown (hypothesis) quantities exist

$$P(hypothesis \mid data) = \frac{P(data \mid hypothesis)P(hypothesis)}{P(data)}$$

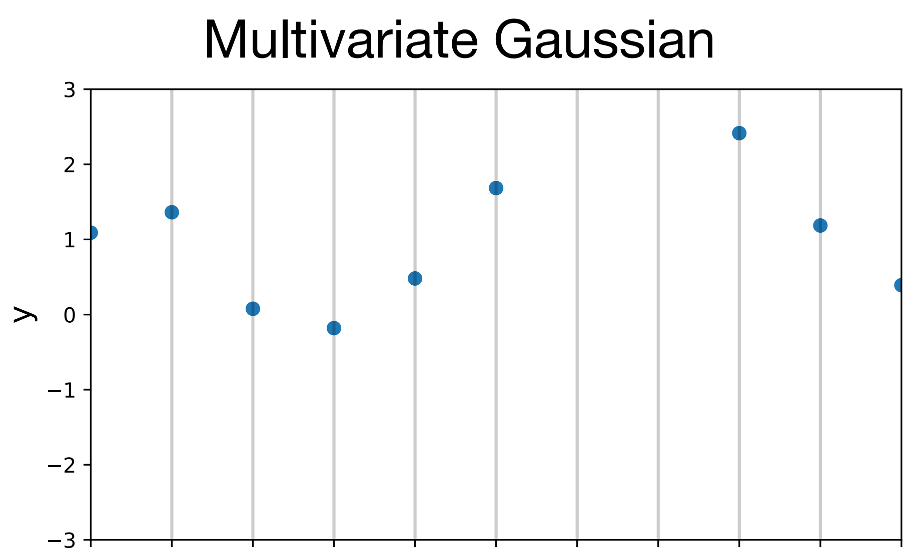
- Inference (Learning)

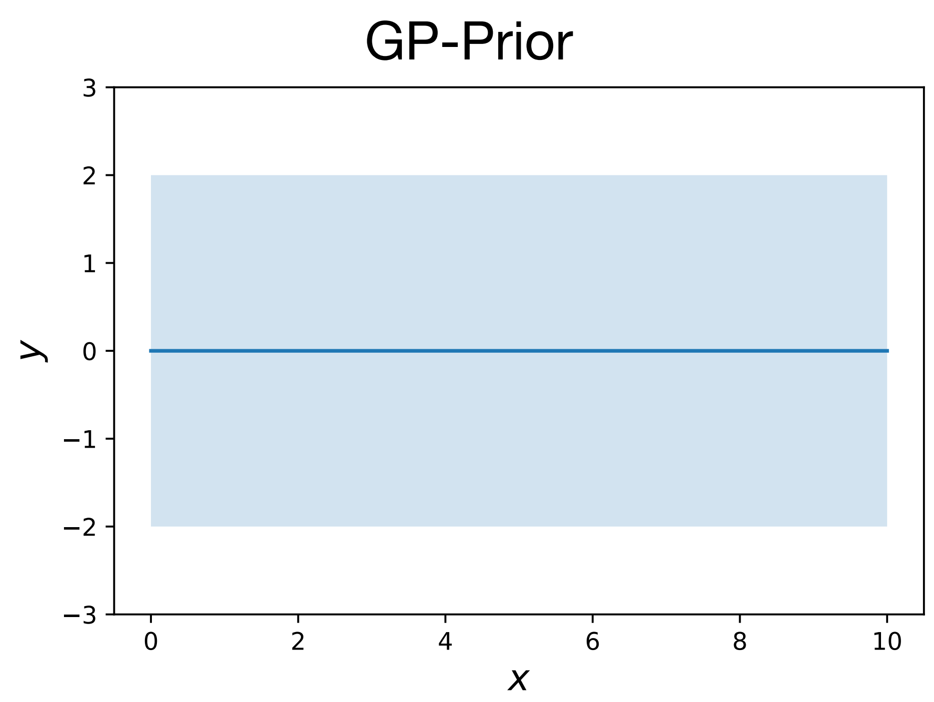
$$P(\theta \mid D) = \frac{P(D \mid \theta)P(\theta)}{P(D)}$$

- Prediction

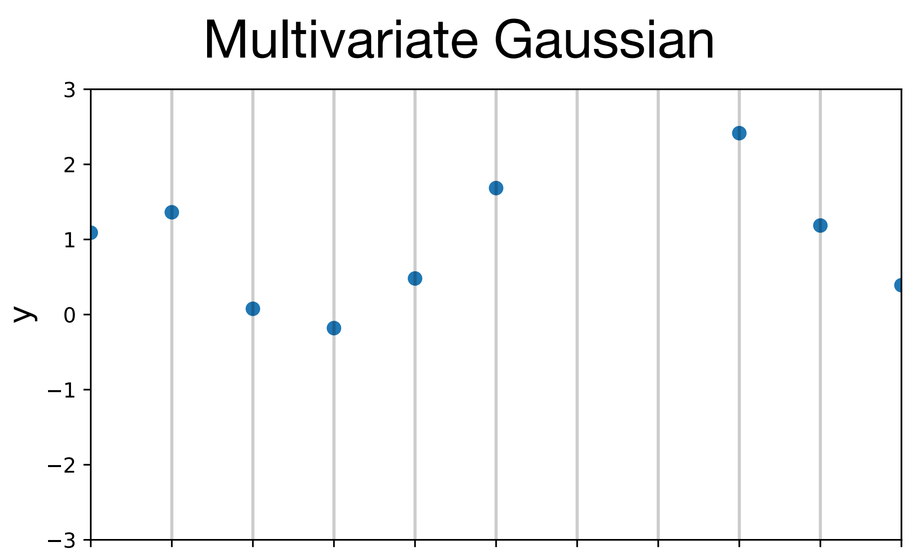
$$P(y \mid D) = \int P(y \mid \theta)P(\theta \mid D)d\theta$$

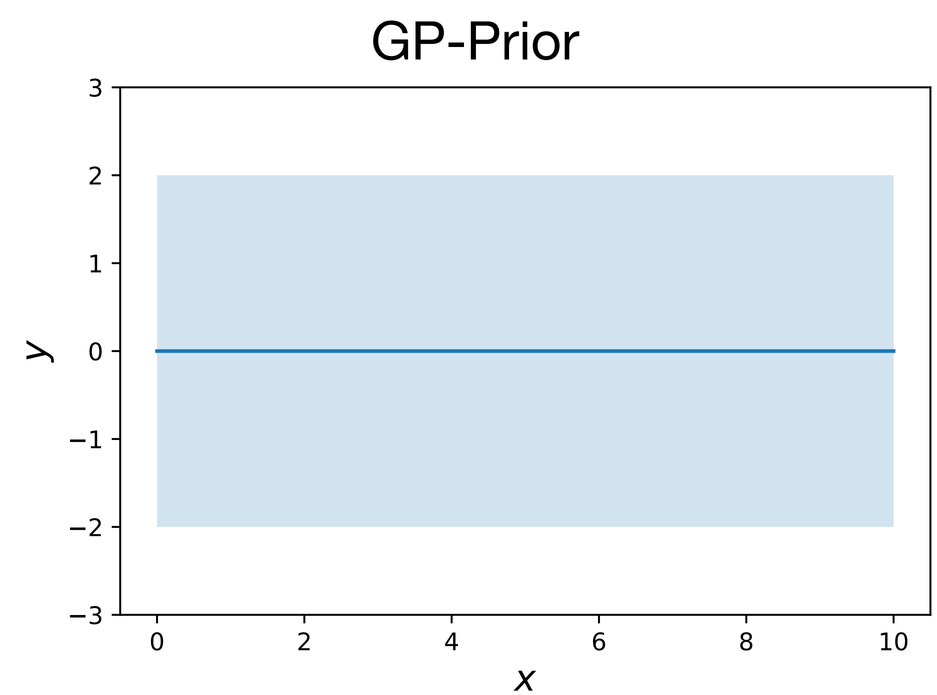
Gaussian Processes

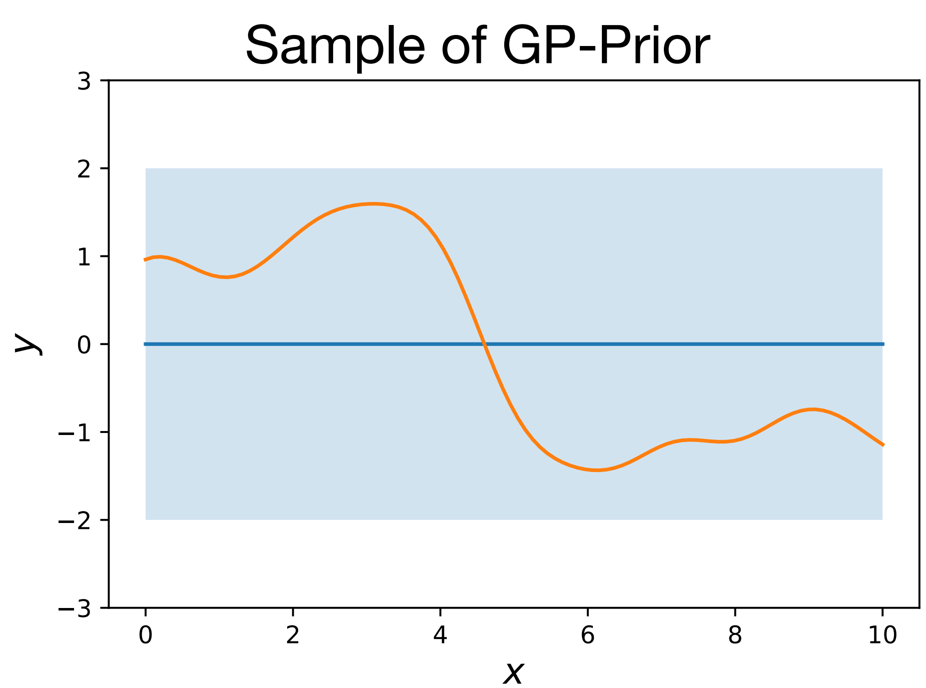




Gaussian Processes





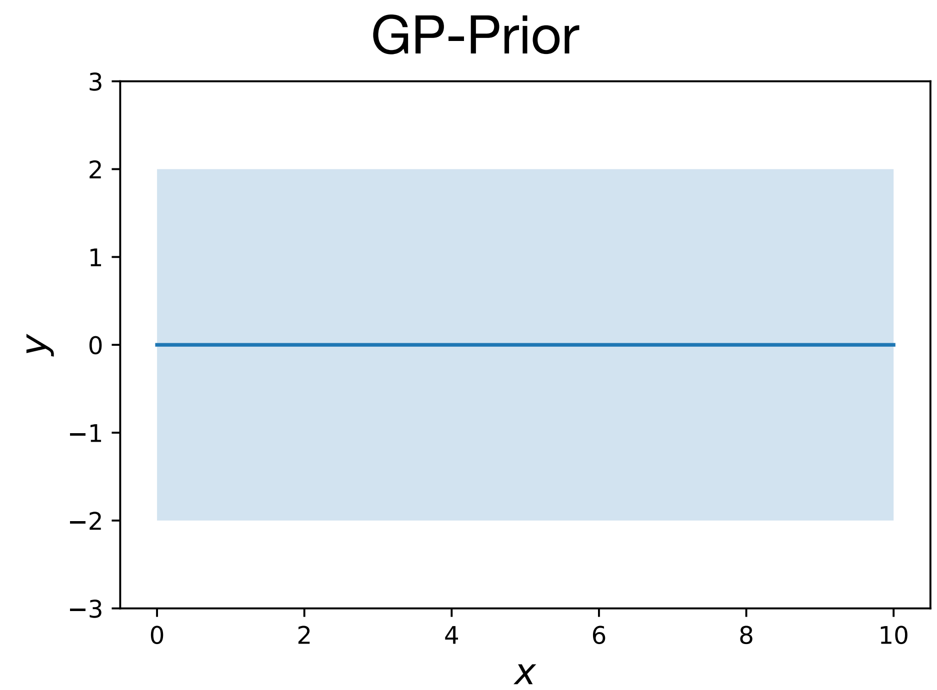


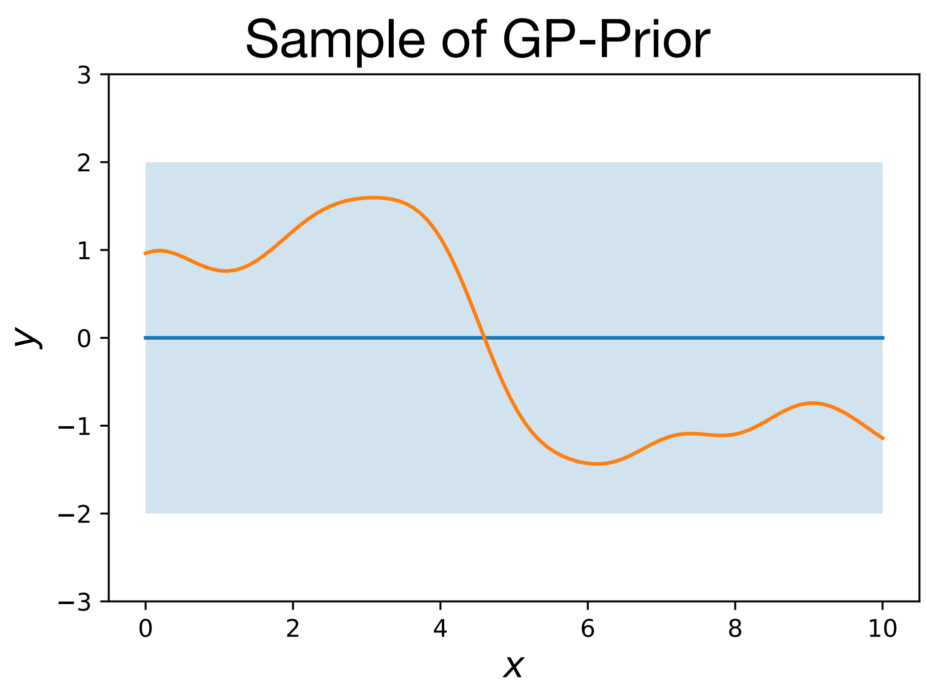
Gaussian Processes

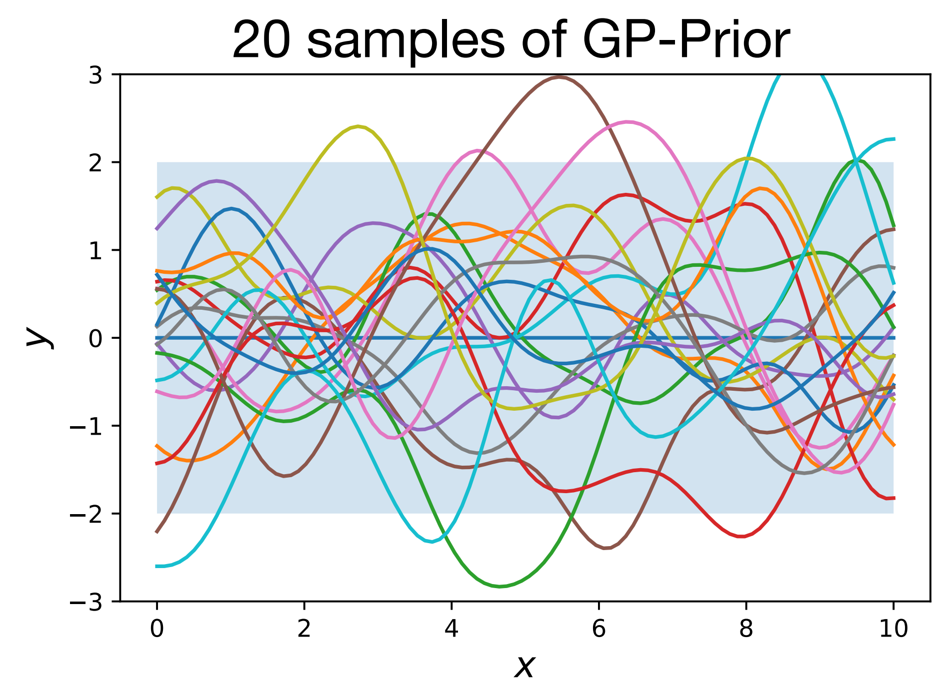
$$Y \sim GP\left(m(x), k(x, x')\right)$$

$$m(x) = 0$$

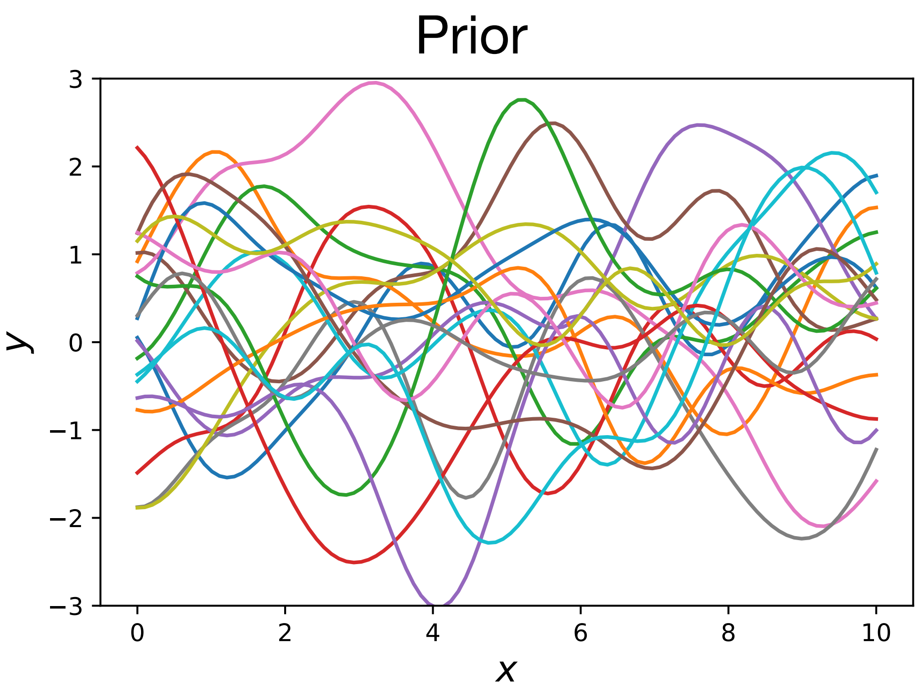
$$k(x, x') = e^{-\frac{1}{2}(x - x')^2}$$



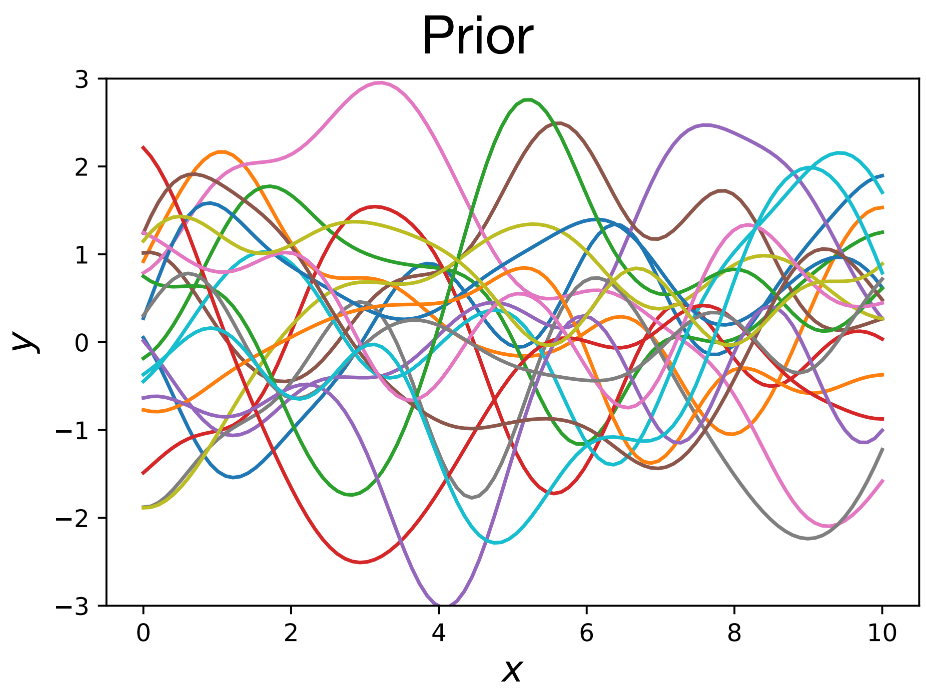


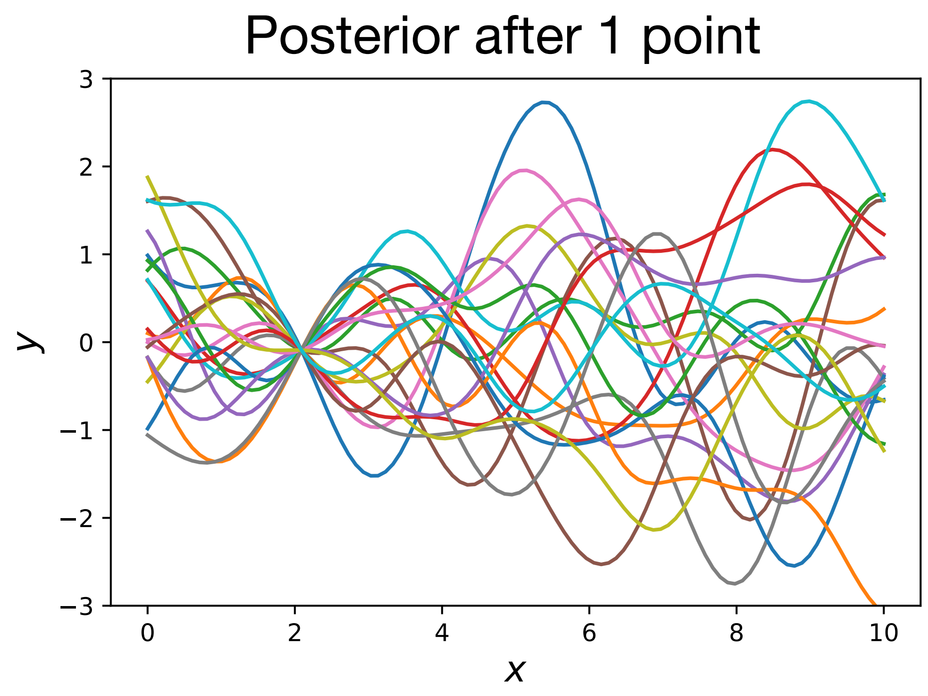


Gaussian Processes

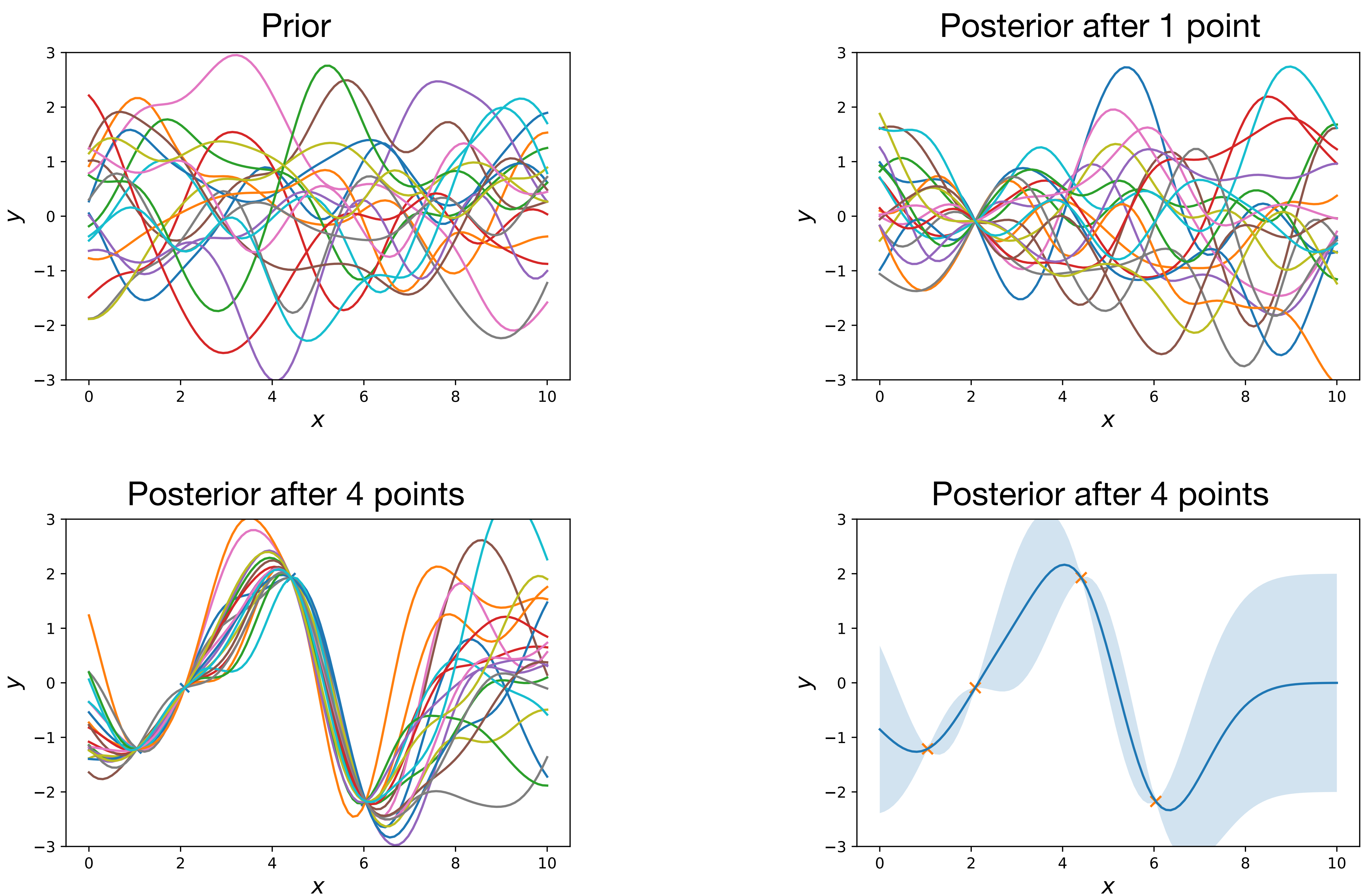


Gaussian Processes





Gaussian Processes

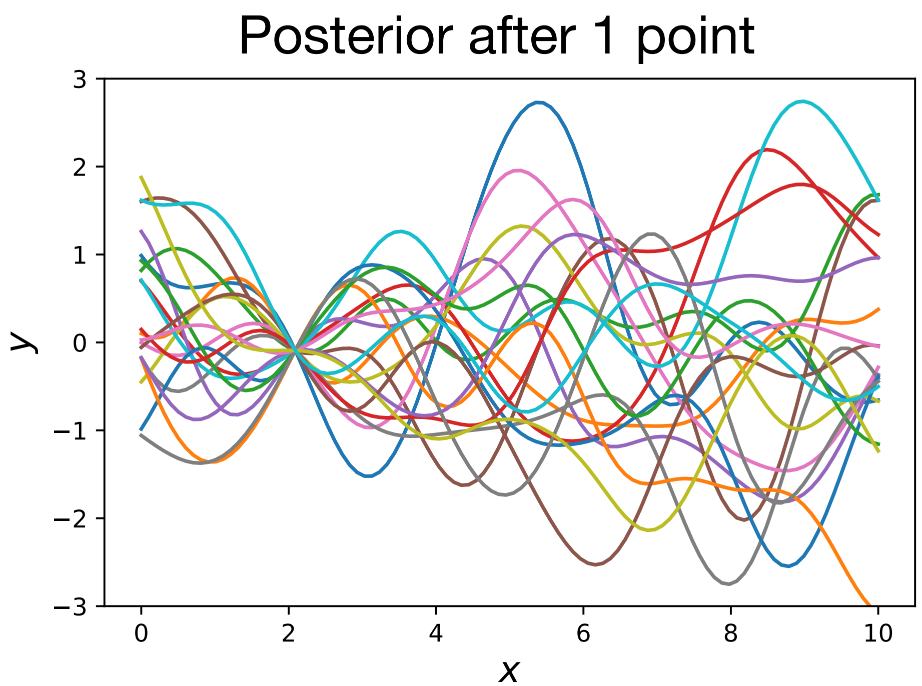


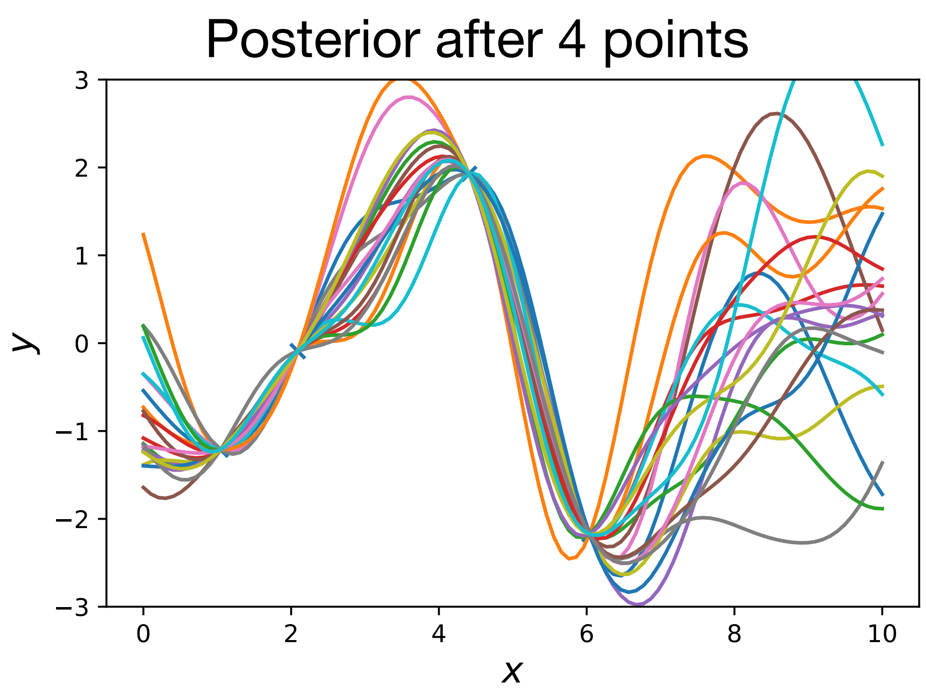
Gaussian Processes

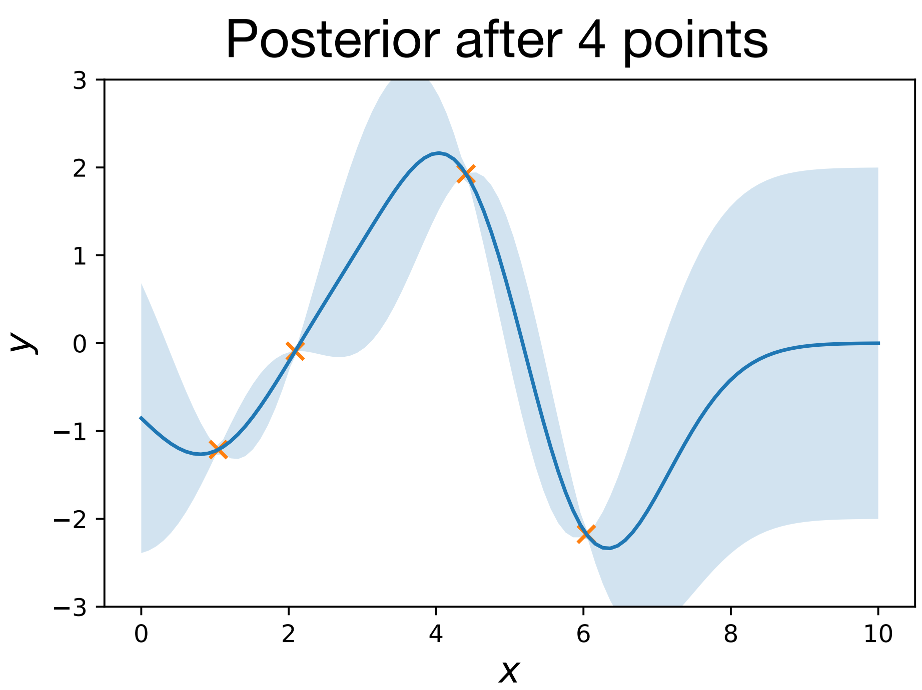
$$Y_* \mid Y, X, X_* \sim \mathcal{N}(m, K)$$

$$m = K(X_*, X)K(X, X)^{-1}Y$$

$$K = K(X_*, X_*) - K(X_*, X)K(X, X)^{-1}K(X, X_*)$$



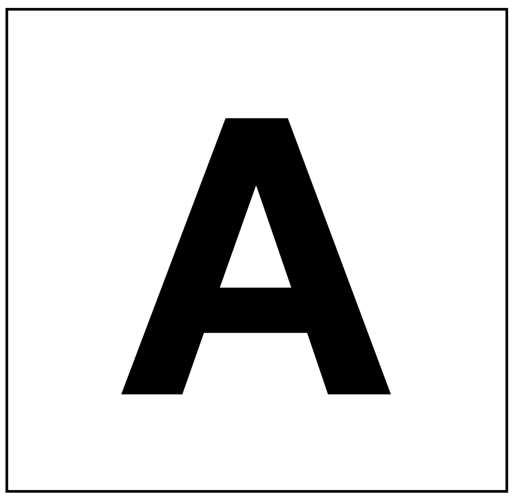




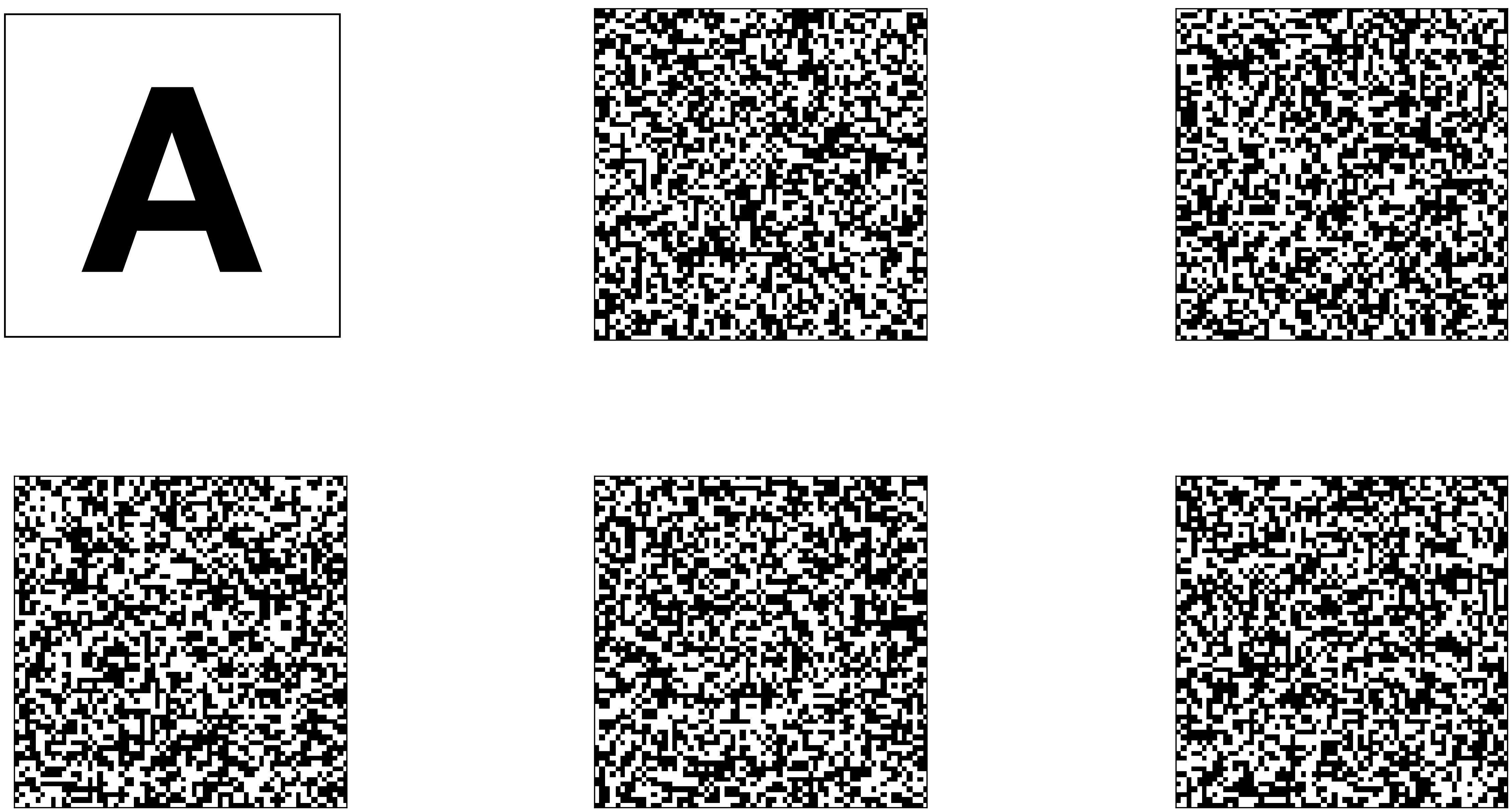
Latent Variable Models

A

Latent Variable Models



Latent Variable Models



Linear Latent Variable Models

Problem: $Y = XW^T + \epsilon$

$$Y \in \mathbb{R}^{N \times D} \qquad N - number\ of\ data\ points$$

$$X \in \mathbb{R}^{N \times Q} \qquad D - dimension\ of\ data\ space$$

$$W \in \mathbb{R}^{Q \times D} \qquad Q - dimension\ of\ latent\ space$$

Linear Latent Variable Models

Problem: $Y = XW^T + \epsilon$

$$Y \in \mathbb{R}^{N\times D} \qquad N - number\ of\ data\ points$$
$$X \in \mathbb{R}^{N\times Q} \qquad D - dimension\ of\ data\ space$$
$$W \in \mathbb{R}^{Q\times D} \qquad Q - dimension\ of\ latent\ space$$

- Principle Component Analysis (PPCA)

$$y_{n,\cdot} = Wx_{n,\cdot} + \epsilon_{n,\cdot}$$

$$X \sim \mathcal{N}(0, \mathrm{I})$$
$$\epsilon \sim \mathcal{N}(0, \sigma^2\mathrm{I})$$

$$p(Y \mid W) = \prod_n \mathcal{N}(y_{n,\cdot} \mid 0, WW^T + \sigma^2\mathrm{I})$$

Linear Latent Variable Models

Problem: $Y = XW^T + \epsilon$

$$Y \in \mathbb{R}^{N \times D} \quad N - number\ of\ data\ points$$
$$X \in \mathbb{R}^{N \times Q} \quad D - dimension\ of\ data\ space$$
$$W \in \mathbb{R}^{Q \times D} \quad Q - dimension\ of\ latent\ space$$

- Principle Component Analysis (PPCA)

$$y_{n,\cdot} = Wx_{n,\cdot} + \epsilon_{n,\cdot}$$

$$X \sim \mathcal{N}(0, \mathrm{I})$$
$$\epsilon \sim \mathcal{N}(0, \sigma^2 \mathrm{I})$$

$$p(Y \mid W) = \prod_n \mathcal{N}(y_{n,\bullet} \mid 0, WW^T + \sigma^2 \mathrm{I})$$

- Dual Principle Component Analysis (Dual PPCA)

$$y_{\bullet,d} = Xw_{\bullet,d} + \epsilon_{\bullet,d}$$

$$W \sim \mathcal{N}(0, \mathrm{I})$$
$$\epsilon \sim \mathcal{N}(0, \sigma^2 \mathrm{I})$$

$$p(Y \mid W) = \prod_d \mathcal{N}(y_{\bullet,d} \mid 0, XX^T + \sigma^2 \mathrm{I})$$

Linear Latent Variable Models

Problem: $Y = XW^T + \epsilon$

$$Y \in \mathbb{R}^{N \times D} \qquad N - \textit{number of data points}$$
$$X \in \mathbb{R}^{N \times Q} \qquad D - \textit{dimension of data space}$$
$$W \in \mathbb{R}^{Q \times D} \qquad Q - \textit{dimension of latent space}$$

- Principle Component Analysis (PPCA)

$$y_{n,\cdot} = Wx_{n,\cdot} + \epsilon_{n,\cdot}$$

$$X \sim \mathcal{N}(0, \mathrm{I})$$
$$\epsilon \sim \mathcal{N}(0, \sigma^2 \mathrm{I})$$

$$p(Y \mid W) = \prod_n \mathcal{N}(y_{n,\bullet} \mid 0, WW^T + \sigma^2 \mathrm{I})$$

- Dual Principle Component Analysis (Dual PPCA)

$$y_{\bullet,d} = Xw_{\bullet,d} + \epsilon_{\bullet,d}$$

$$W \sim \mathcal{N}(0, \mathrm{I})$$
$$\epsilon \sim \mathcal{N}(0, \sigma^2 \mathrm{I})$$

$$p(Y \mid W) = \prod_d \mathcal{N}(y_{\bullet,d} \mid 0, XX^T + \sigma^2 \mathrm{I})$$

$$(XX^T)_{ij} = X_i^T X_j$$

Linear Latent Variable Models

Problem: $Y = XW^T + \epsilon$

$$Y \in \mathbb{R}^{N \times D} \qquad N - \textit{number of data points}$$
$$X \in \mathbb{R}^{N \times Q} \qquad D - \textit{dimension of data space}$$
$$W \in \mathbb{R}^{Q \times D} \qquad Q - \textit{dimension of latent space}$$

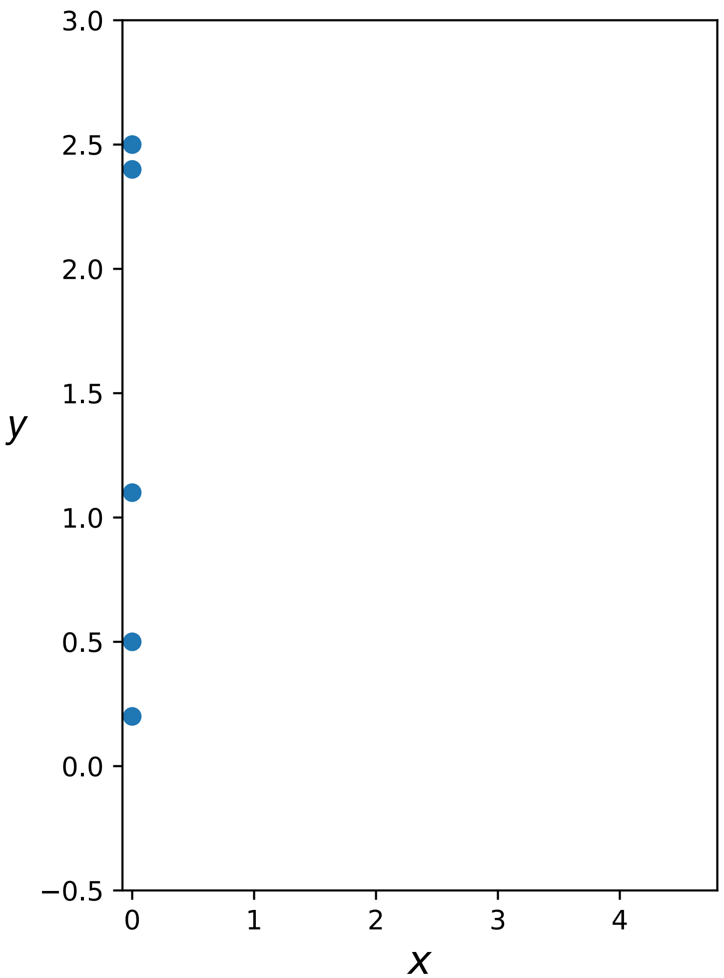
- Principle Component Analysis (PPCA)

$$y_{n,\cdot} = Wx_{n,\cdot} + \epsilon_{n,\cdot}$$
$$X \sim \mathcal{N}(0, \mathrm{I})$$
$$\epsilon \sim \mathcal{N}(0, \sigma^2 \mathrm{I})$$
$$p(Y \mid W) = \prod_n \mathcal{N}(y_{n,\bullet} \mid 0, WW^T + \sigma^2 \mathrm{I})$$

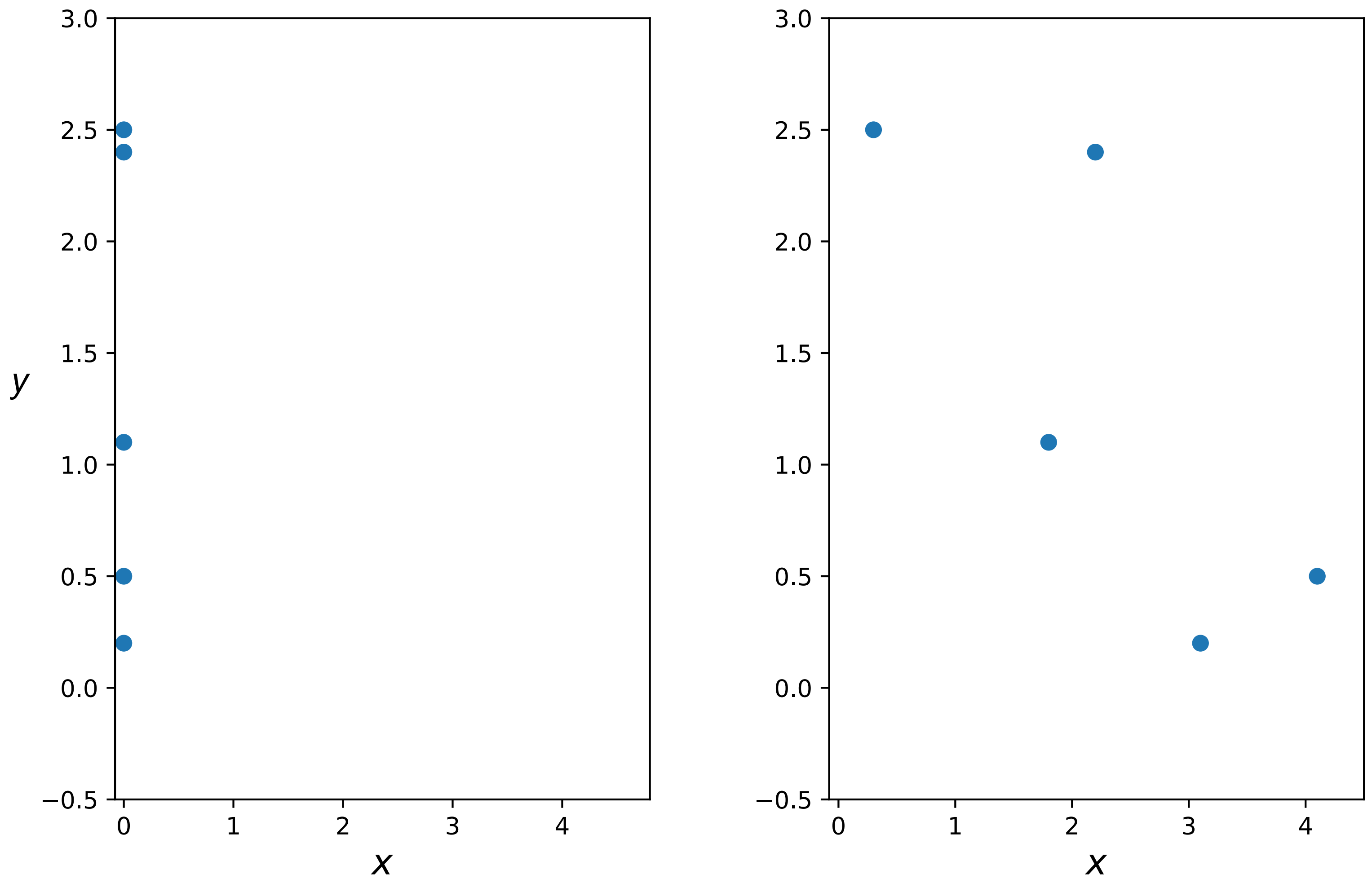
- Dual Principle Component Analysis (Dual PPCA)

$$y_{\bullet,d} = Xw_{\bullet,d} + \epsilon_{\bullet,d}$$
$$W \sim \mathcal{N}(0, \mathrm{I})$$
$$\epsilon \sim \mathcal{N}(0, \sigma^2 \mathrm{I})$$
$$p(Y \mid W) = \prod_d \mathcal{N}(y_{\bullet,d} \mid 0, XX^T + \sigma^2 \mathrm{I})$$
$$(XX^T)_{ij} = X_i^T X_j = k(X_i, X_j)$$

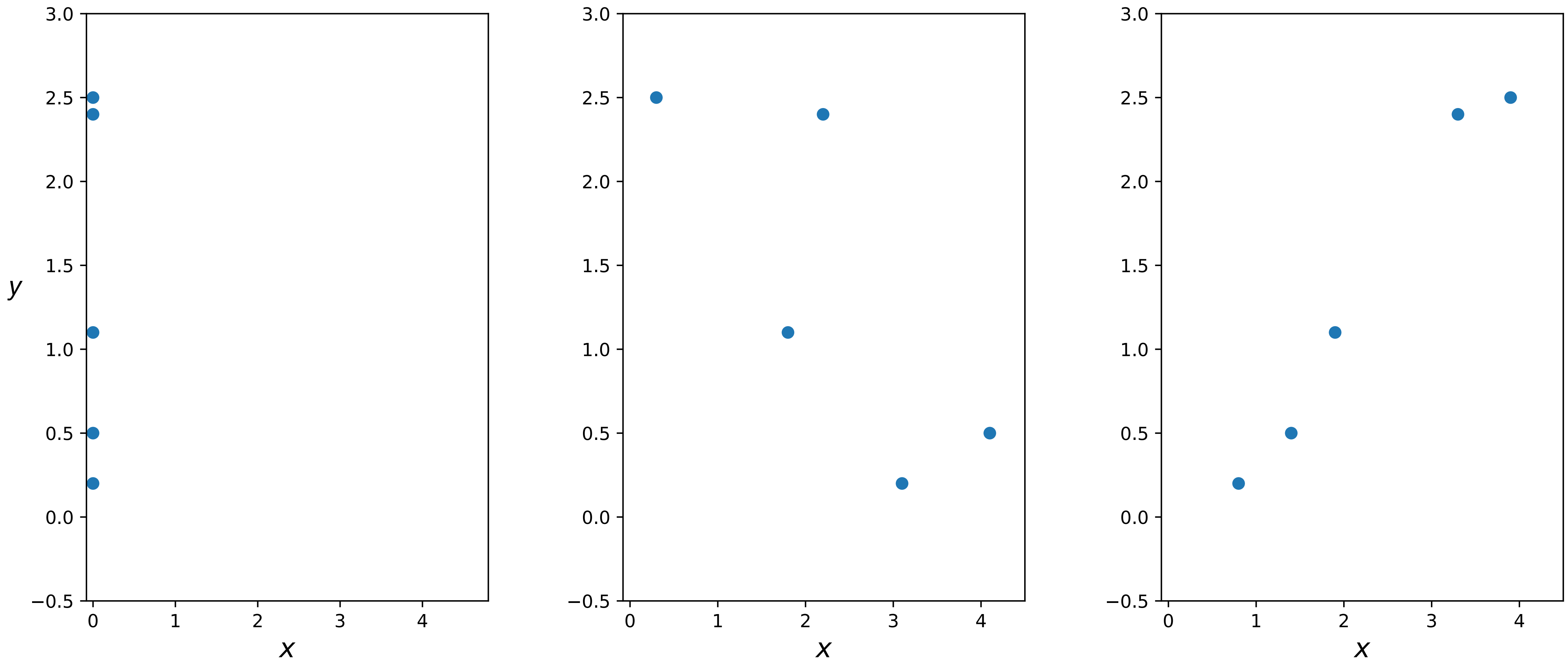
Gaussian Process Latent Variable Model



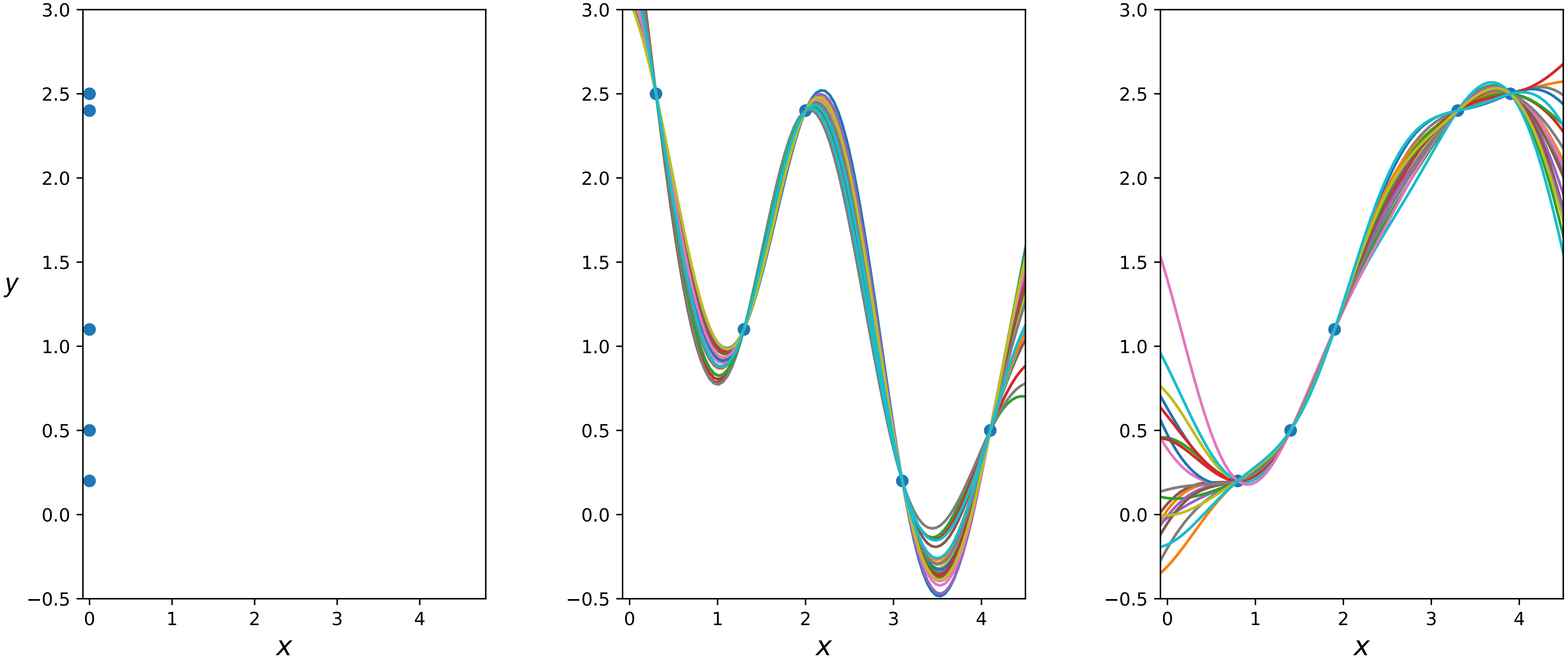
Gaussian Process Latent Variable Model



Gaussian Process Latent Variable Model



Gaussian Process Latent Variable Model



Applications in Finance

- CAPM $r_n - r_f = \alpha_n + \beta_n (r_m - r_f) + \epsilon$
- In an efficient market $\mathbb{E}[\alpha_i] = 0$

$$\widetilde{r_n} = \beta_n \widetilde{r_m} + \epsilon$$

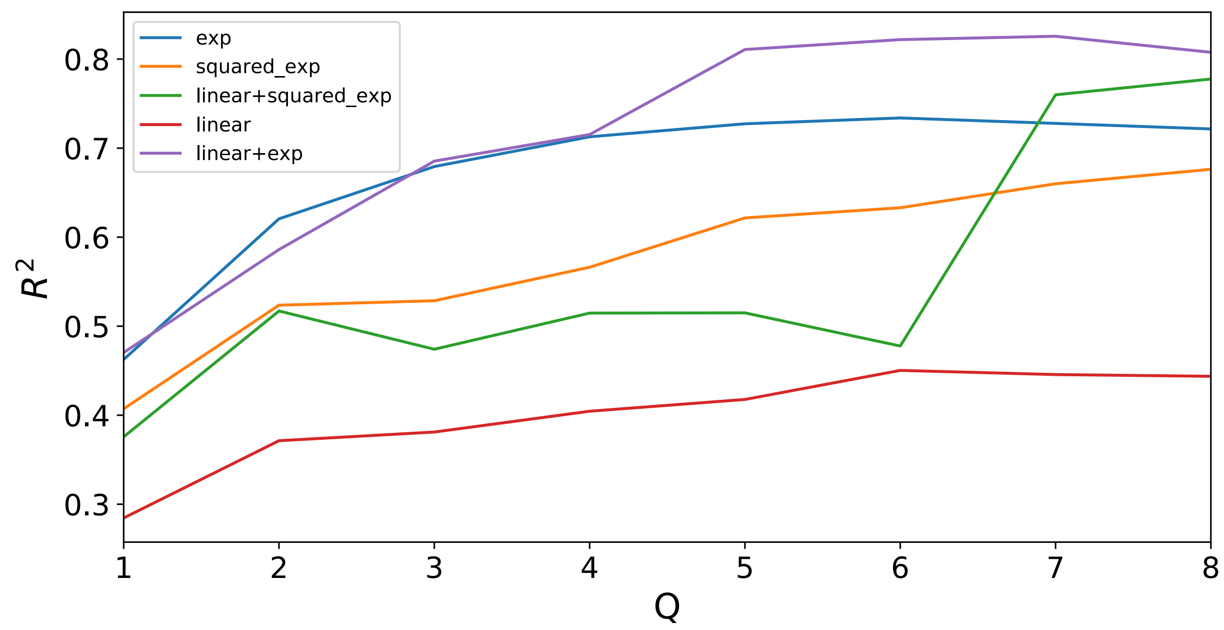
- Given N different stocks on D days we denote the return matrix

$$R = (r_1, \ldots, r_N)^T \in \mathbb{R}^{N \times D}$$

- We can solve this model with GP-LVM and learn the covariance between stocks
- Markowitz portfolio theory $w_{opt} = \min_w (w^T K w - q \mu^T w)$

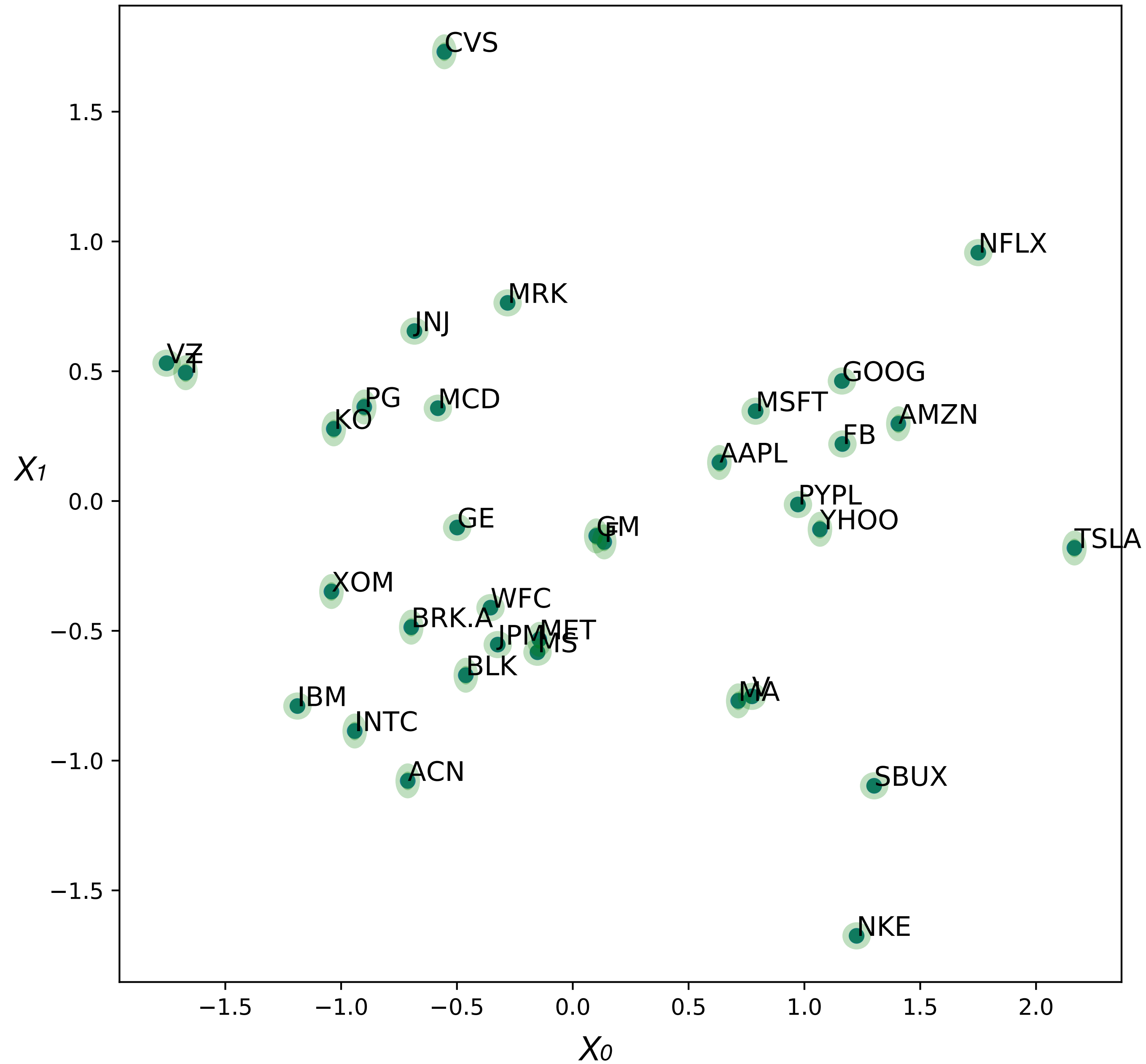
Applications in Finance

R squared:

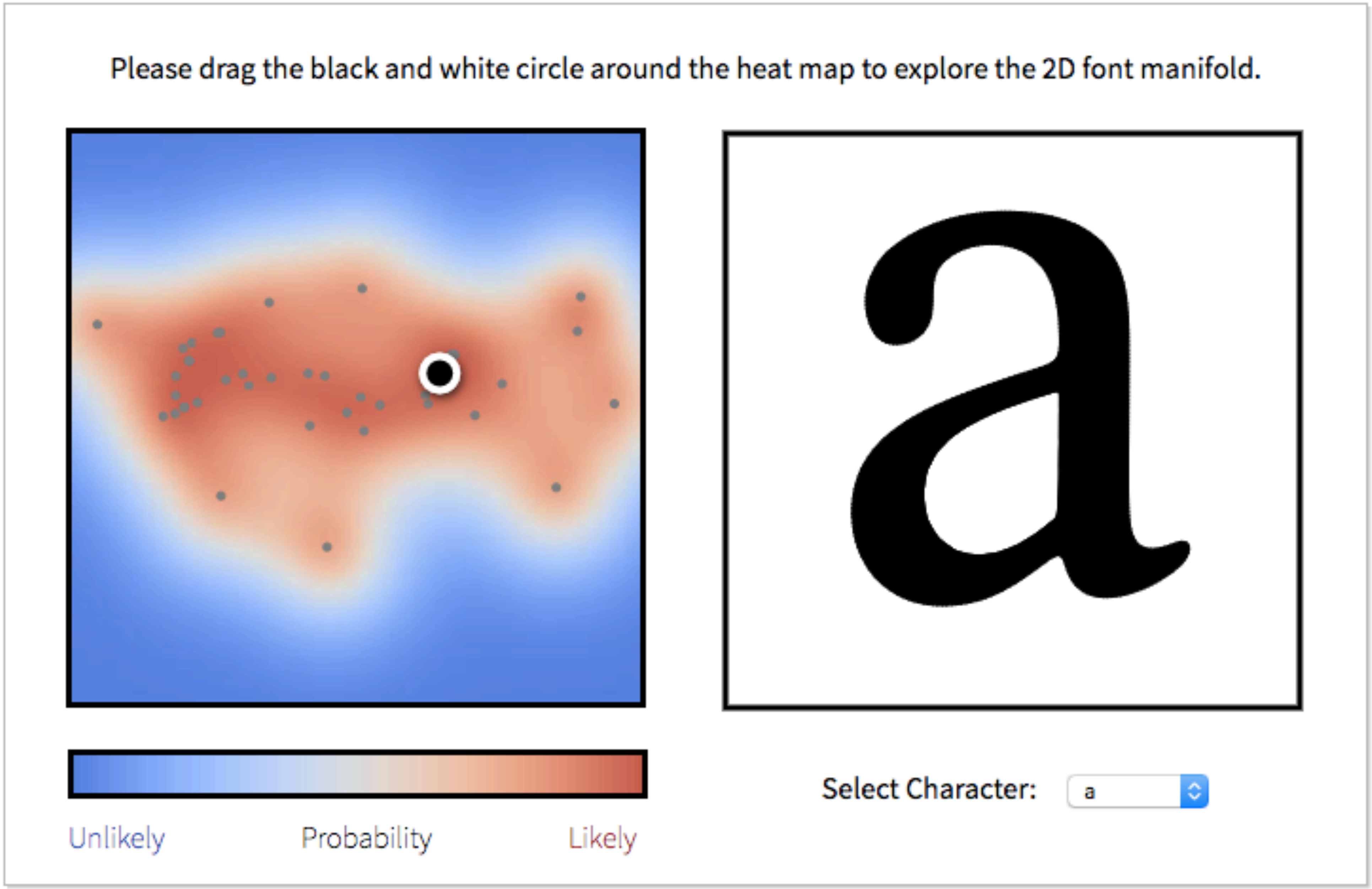


Returns for learning period of 1 year and prediction period of 6 months

Applications in Finance



Learning a Manifold of Fonts



Learning a Manifold of Fonts. Campbell (2014)

Thanks for your attention!

Supervisor: Prof. Dr. Nils Bertschinger
Funder: Dr. h. c. Helmut O. Maucher