

ガウス過程の基礎

統計数理研究所

松井知子

目次

- ガウス過程 (Gaussian Process; GP)
 - 序論
 - GPによる回帰
 - GPによる識別
- GP状態空間モデル
 - 概括
 - GP状態空間モデルによる音楽ムードの推定

GP序論

- ノンパラメトリック予測
- カーネル法の利用
- 参考文献:
 - C. E. Rasmussen and C. K. I. Williams
「Gaussian Processes for Machine Learning」
 - David Barber 「Gaussian Process (chapter 19 in Bayesian Reasoning and Machine Learning)」
- GP関連サイト
 - <http://www.gaussianprocess.org>
 - <http://www.gaussianprocess.org/gpml/>

GP序論

E. Ebden (A quick introduction)

Typical prediction problem: given some inputs x and the corresponding noisy outputs y , the best estimate y_* for a new input x_* is?

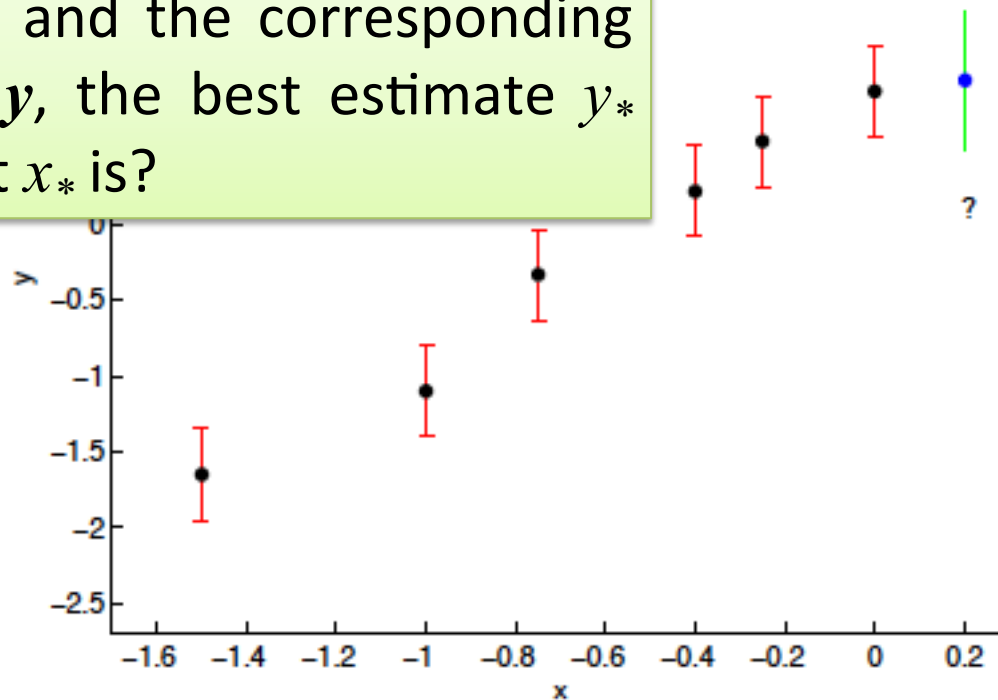


Figure 1: Given six noisy data points (error bars are indicated with vertical lines), we are interested in estimating a seventh at $x_* = 0.2$.

GP序論： $y = f(x)$?

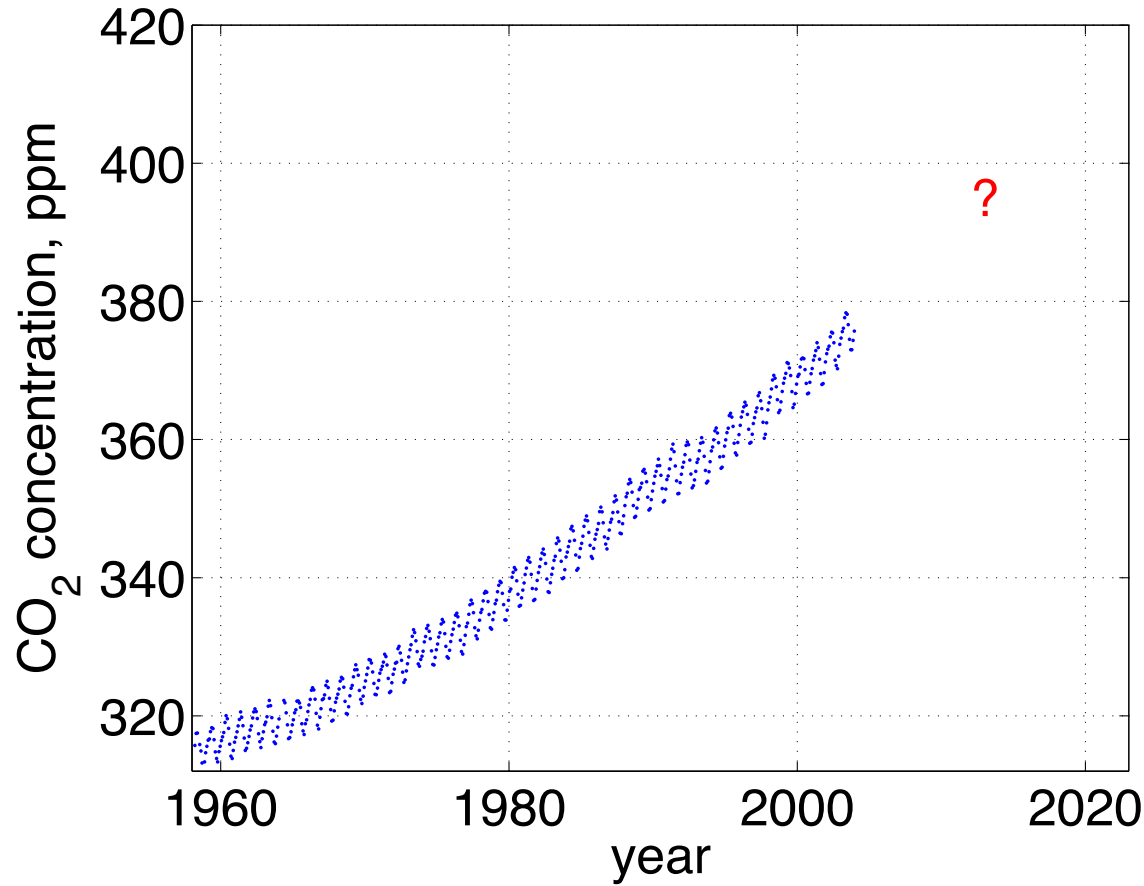
- 線形モデルを仮定
 - ➡ 最小二乗法による線形回帰

$$y = \sum_i w_i \phi_i(x)$$

- 2次、3次、非多項式モデルなどから選択
 - ➡ モデル選択法を利用
- ガウス過程で表す
 - ➡ データに基づいてモデルを自動決定

GP序論

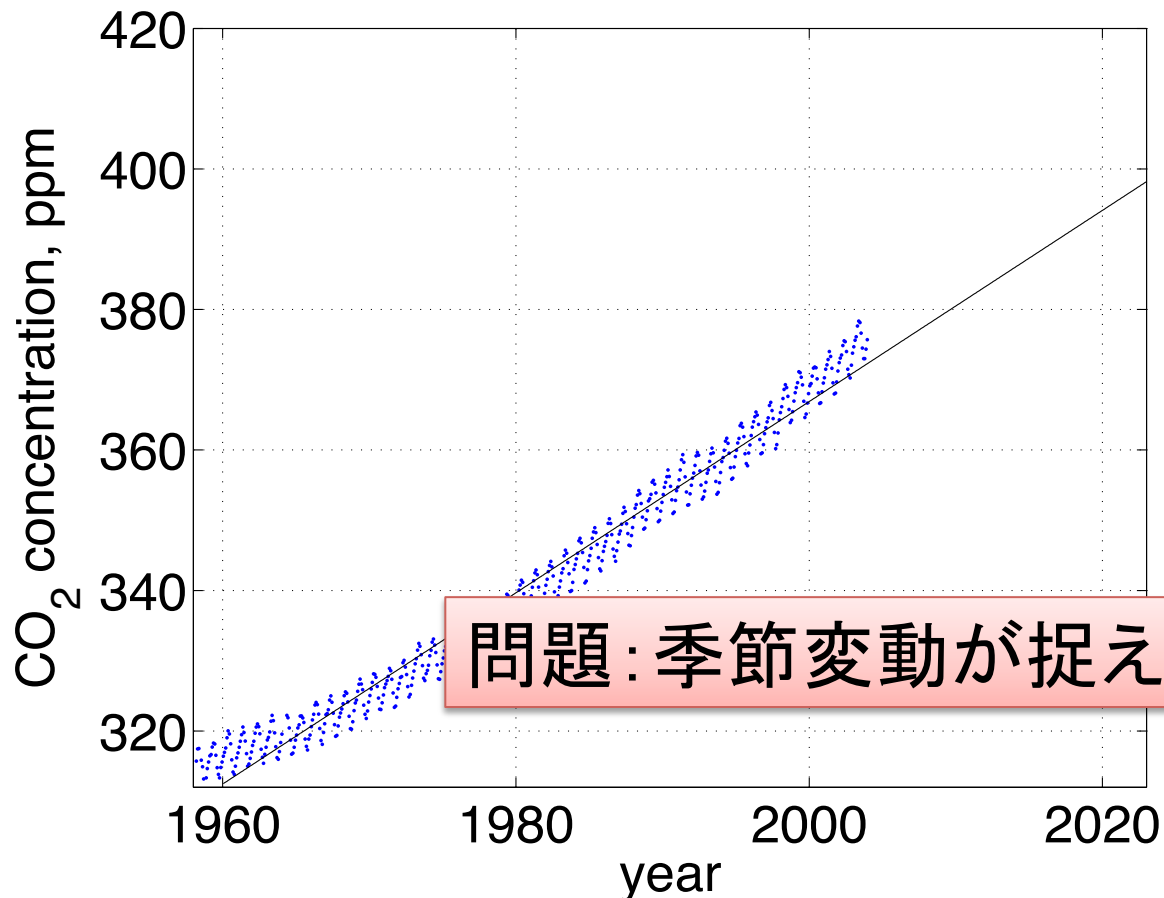
C. E. Rasmussen Example – CO₂ per year



GP序論

C. E. Rasmussen Example – CO₂ per year

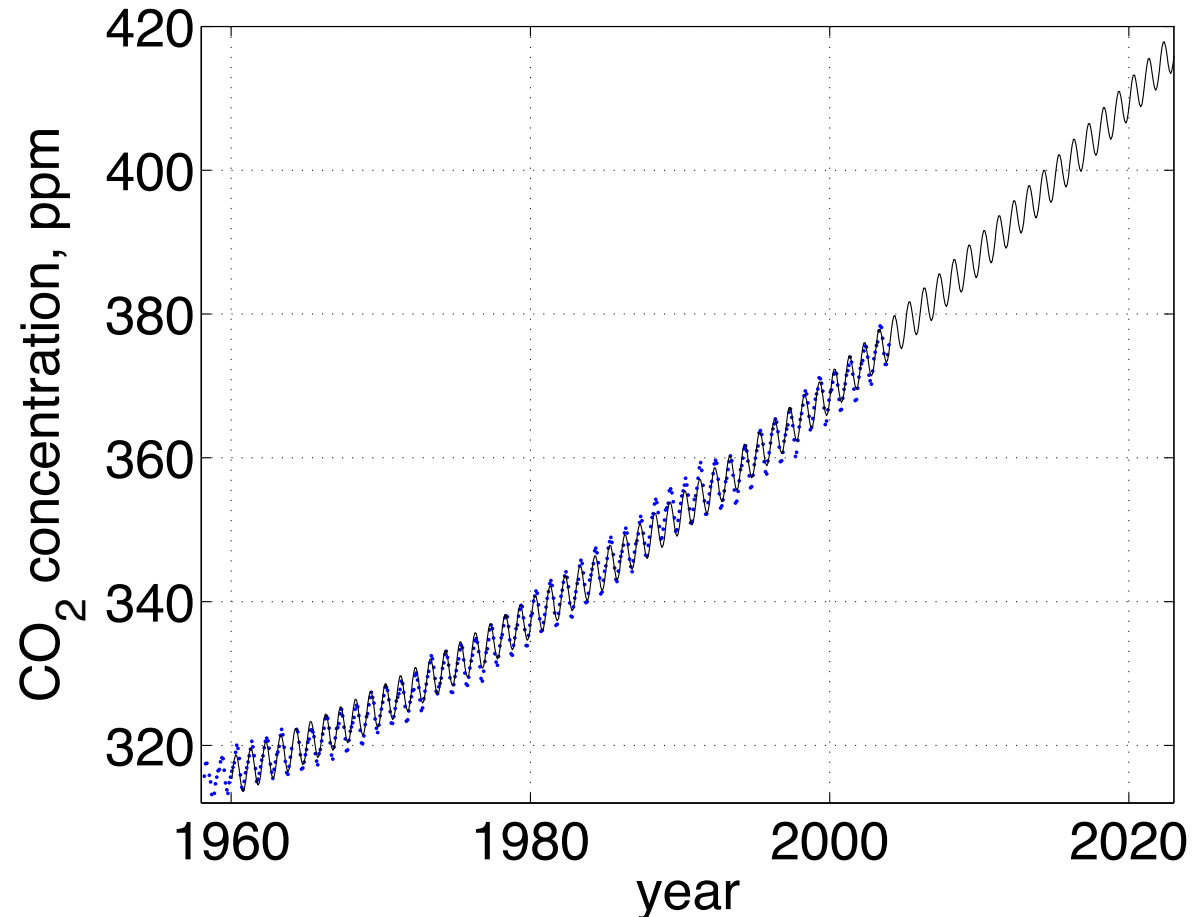
- 最小2乗法による線形モデルの当てはめ



GP序論

C. E. Rasmussen Example – CO₂ per year

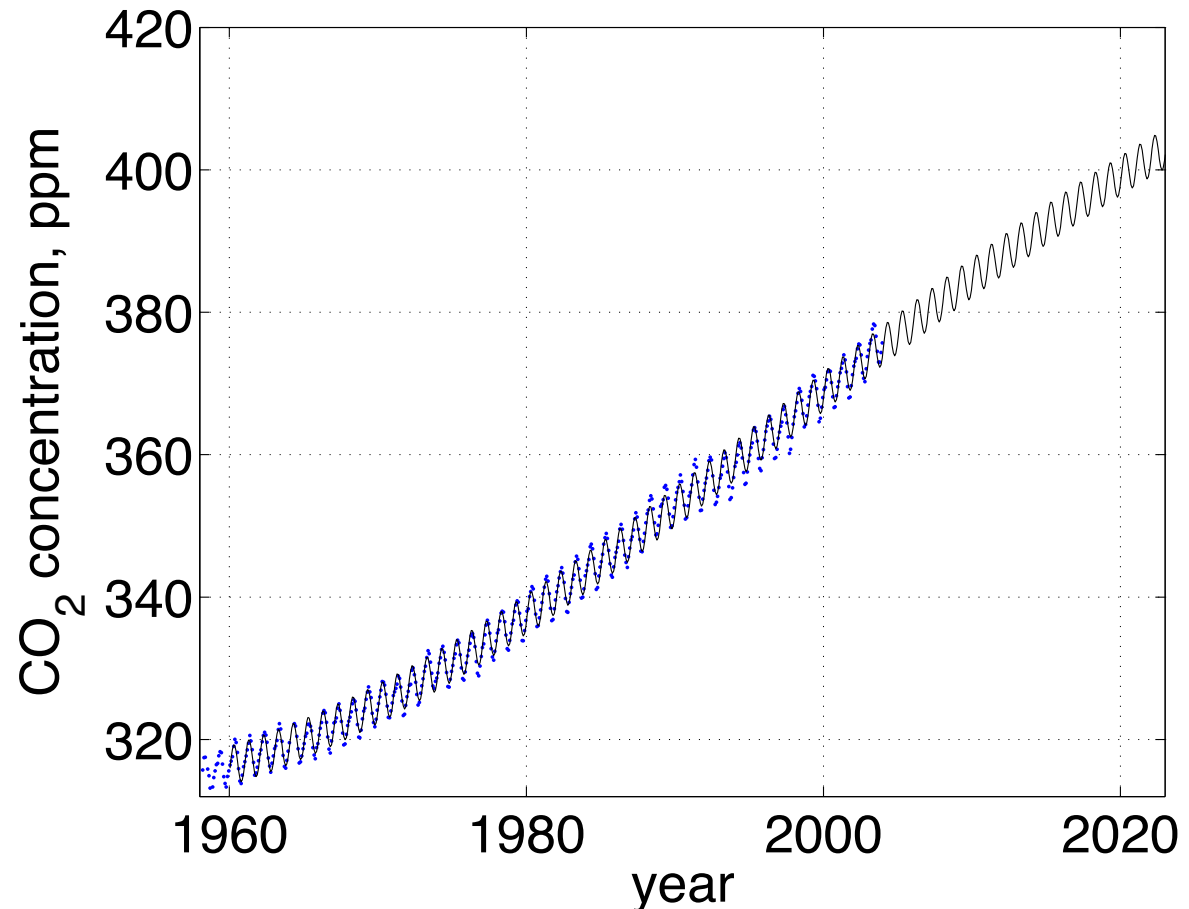
- {季節変動-正弦関数} + {トレンド-**3次**多項式}モデルの当てはめ



GP序論

C. E. Rasmussen Example – CO₂ per year

- {季節変動-正弦関数} + {トレンド-2次多項式}モデルの当てはめ



GP序論

C. E. Rasmussen Example – CO2 per year

- {季節変動-正弦関数} + {トレンド-多項式} 回帰モデル:
 - ✓ データがある部分についてはよく適合している。
 - ✓ データがない部分については、トレンド成分を表す基底関数を3次から2次の多項式に変更しただけで、予測が異なる。

問題: パラメトリックなモデルの利用の難しさ

- 将来のCO2はどのように予測したらよいか？

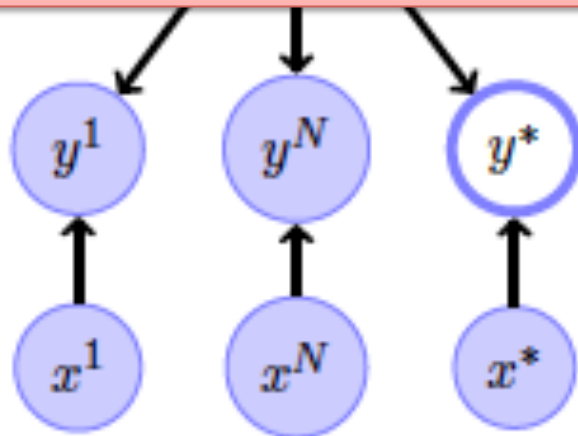


GPの利用

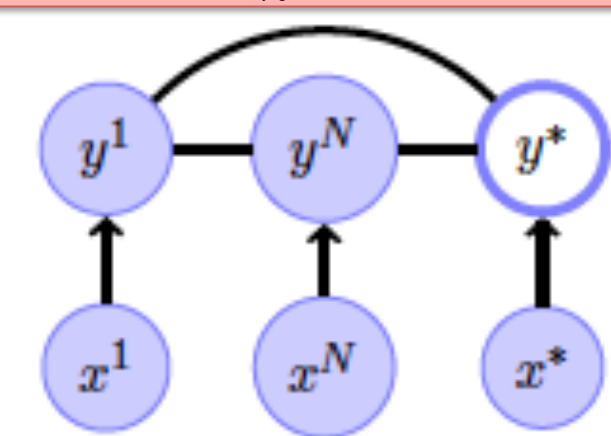
GP序論：ノンパラメトリック予測

- n 番目の入力: x^n 、出力: y^n
- 学習データ: $\mathcal{D} = \{(x^n, y^n), n = 1, \dots, N\} = \mathcal{X} \cup \mathcal{Y}$

$$p(y^*, \mathcal{Y} | x^*, \mathcal{X}) = \int_{\theta} p(y^* | x^*, \theta) p(\theta) \prod_n p(y^n | \theta, x^n)$$



パラメトリックモデル



ノンパラメトリックモデル

GP序論：線形モデルからGPへ

- 線形モデル

$$\mathbf{y} = \Phi \mathbf{w}, \quad \Phi = [\phi(x_1), \dots, \phi(x_N)]^T, \quad p(\mathbf{w}) = \mathcal{N}(\mathbf{w} | 0, \Sigma_{\mathbf{w}})$$

$p(\mathbf{y} | \mathbf{x}) = \int \delta(\mathbf{y} - \Phi \mathbf{w}) p(\mathbf{w}) : \text{Gaussian distributed with}$

$$\text{mean: } \langle \mathbf{y} \rangle = \Phi \langle \mathbf{w} \rangle_{p(\mathbf{w})} = 0$$

$$\text{cov: } \langle \mathbf{y} \mathbf{y}^T \rangle = \Phi \langle \mathbf{w} \mathbf{w}^T \rangle_{p(\mathbf{w})} \Phi^T = \Phi \Sigma_{\mathbf{w}} \Phi^T = \mathbf{K}$$

$\Sigma_{\mathbf{w}} = \mathbf{I} \rightarrow \Phi \Phi^T$

GP序論：線形モデルからGPへ

- 線形モデル

$$\mathbf{y} = \Phi \mathbf{w}, \quad \Phi = [\phi(x_1), \dots, \phi(x_N)]^T, \quad p(\mathbf{w}) = \mathbf{N}(\mathbf{w} | 0, \Sigma_{\mathbf{w}})$$

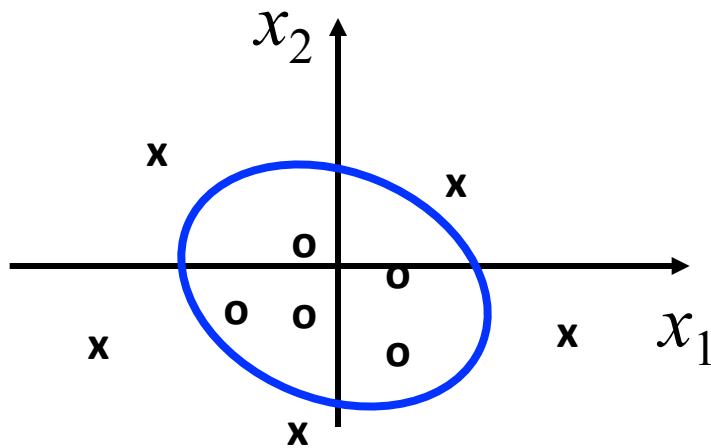


$$p(\mathbf{y} | \mathbf{x}) = \mathbf{N}(\mathbf{y} | 0, \mathbf{K})$$

$$\text{Gram matrix: } [\mathbf{K}]_{n,n'} = \phi(x_n)^T \phi(x_{n'}) = k(x_n, x_{n'})$$

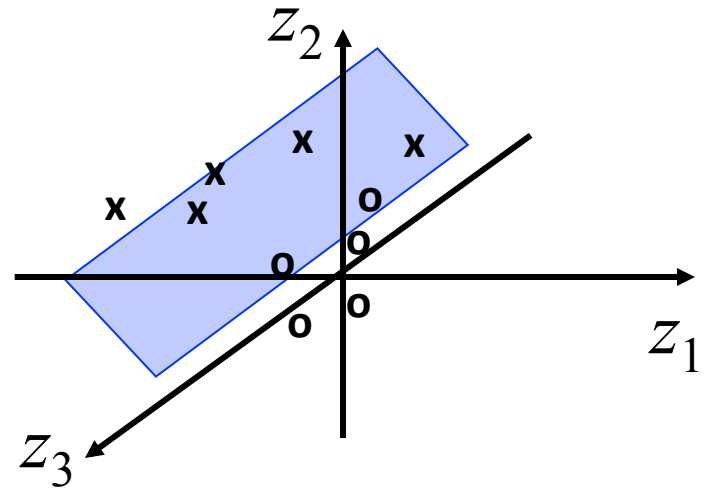
GP: ノンパラメトリックモデル

GP序論:カーネル法



$$(x_1, x_2)$$

入力空間 R^2



$$(z_1, z_2, z_3) := (x_1^2, x_2^2, x_1x_2)$$

特徴空間 R^3

GP序論

カーネル法：高次元空間での内積計算

- 2次元から3次元空間への写像

$$\mathbf{x} = (x_1, x_2)$$

$$\Phi(\mathbf{x}) = (x_1^2, \sqrt{2}x_1x_2, x_2^2)$$

- カーネル関数

$$K(\mathbf{x}, \mathbf{y}) := \langle \mathbf{x}, \mathbf{y} \rangle^2$$

$$\langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle = \langle (x_1^2, \sqrt{2}x_1x_2, x_2^2), (y_1^2, \sqrt{2}y_1y_2, y_2^2) \rangle$$

$$= \langle (x_1, x_2), (y_1, y_2) \rangle^2$$

$$= \langle \mathbf{x}, \mathbf{y} \rangle^2$$

GP序論

カーネル法：次元の呪いの克服

- カーネル関数

$$\mathbf{K}(\mathbf{x}, \mathbf{y}) = (\langle \mathbf{x}, \mathbf{y} \rangle + 1)^s$$

実際の計算！

高々64次元の内積

\mathbf{x} が64次元で、 $s = 3$ の時

$$\mathbf{x} = (x_1, \dots, x_{64})$$

.....▶ [入力空間]

$$\Phi(\mathbf{x}) = (1, x_1, \dots, x_{64},$$

$$x_1^2, \dots, \sqrt{2}x_1x_2, \dots,$$

$$x_1^3, \dots, \sqrt{3}x_1^2x_2, \dots, \sqrt{6}x_1x_2x_3, \dots,$$

$$x_1^4, \dots)$$

.....▶ [特徴空間]

$$\langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle$$

47905次元の内積、大変な計算！

GP序論

カーネル法：識別関数

$$f(\mathbf{x}) = \mathbf{w}^T \Phi(\mathbf{x}) + b$$

$$= (w_1 x_1 x_2 x_3 + \cdots + w_{41664} x_{62} x_{63} x_{64}$$

$$+ w_{41665} x_1^2 x_2 + \cdots + w_{45696} x_{64}^2 x_{63}$$

$$+ w_{45697} x_1 x_2 + \cdots + w_{47712} x_{63} x_{64}$$

$$+ w_{47713} x_1 + \cdots + w_{47776} x_{64}$$

$$+ w_{47777} x_1^2 + \cdots + w_{47840} x_{64}^2$$

$$+ w_{47841} x_1^3 + \cdots + w_{47904} x_{64}^3$$

$$+ w_{47905} \cdot 1) + b = \sum_{i \in SV} \alpha_i y_i \mathbf{k}(\mathbf{x}_i, \mathbf{x}) + b$$

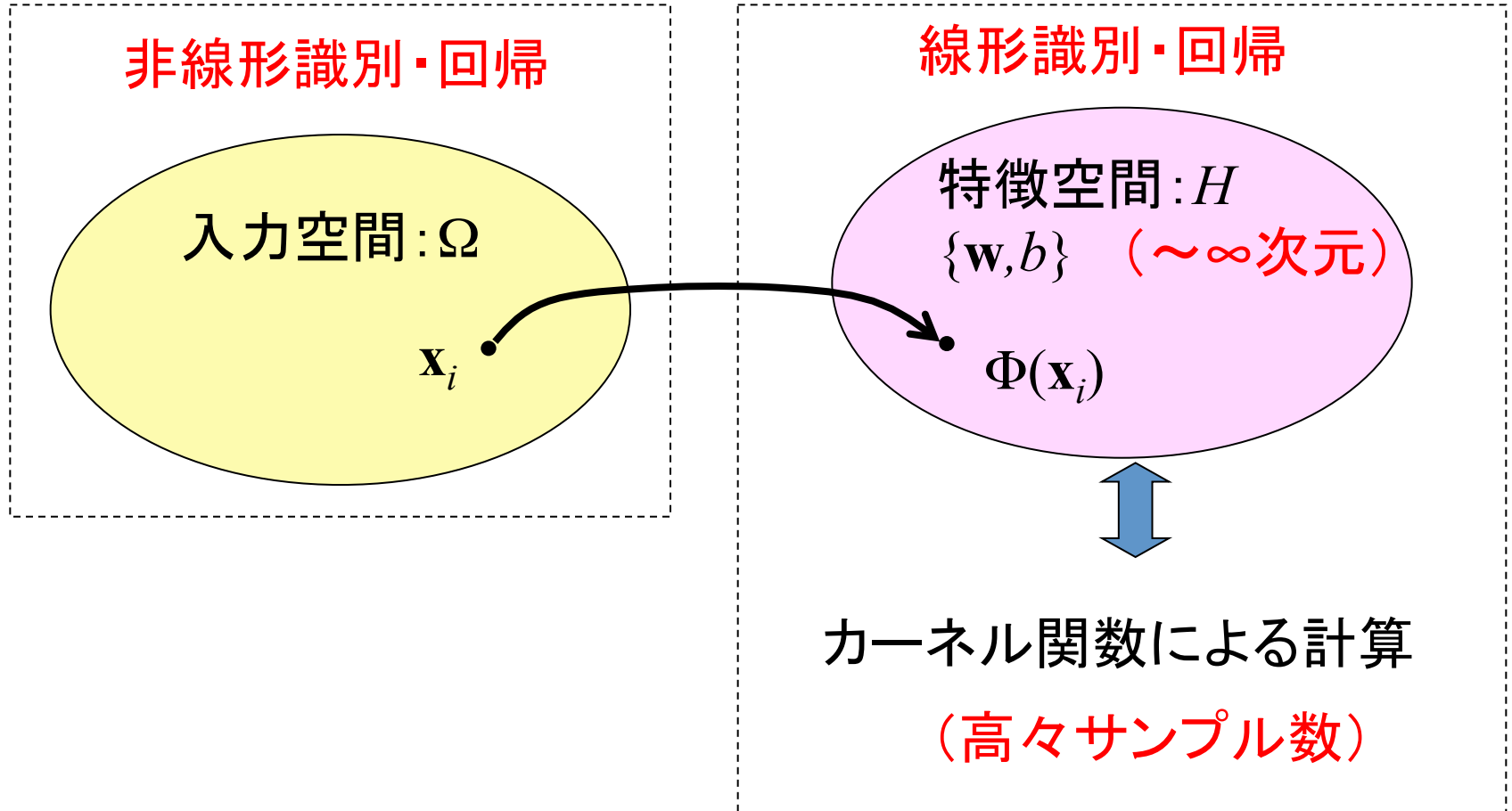
\mathbf{x} : 64次元ベクトル
 $\mathbf{K}(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^T \mathbf{y} + 1)^3$

高い表現力！

$|w_j|$ が大きければ識別に有効な特徴量

GP序論

カーネル法：入力空間と特徴空間



GP序論

カーネル法: $k(x, y) = \Phi^T(x) \Phi(y)$ となる条件

1. 正定値カーネル
2. 再生核ヒルベルト空間

- 集合 Ω , $k: \Omega \times \Omega \rightarrow \mathbf{R}$
- $k(x, y)$: Ω 上の正定値カーネル

[対称性] $k(x, y) = k(y, x)$

[正定値性] 任意の自然数 n と任意の Ω の点 x_1, \dots, x_n に対して、

$n \times n$ グラム行列

$$\{k(x_i, x_j)\}_{i,j=1}^n = \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_n) \\ \vdots & \ddots & \vdots \\ k(x_n, x_1) & \cdots & k(x_n, x_n) \end{bmatrix}$$

が半正定値。すなわち、任意の実数 c_1, \dots, c_n に対して、

$$\sum_{i,j=1}^n c_i c_j k(x_i, x_j) \geq 0$$

GP序論：正定値カーネル

$$\Omega = \mathbf{R}^m$$

- 多項式

$$k(x, x') = (x^T x' + \sigma_0^2)^p$$

- Squared exponential

$$k(x, x') = \exp\left(-\frac{\|x - x'\|^2}{2l^2}\right)$$

- γ -exponential

$$k(x, x') = \exp\left(-\left(\frac{\|x - x'\|}{l}\right)^\gamma\right)$$

GP序論:カーネル設計

- カーネルの特質:
 - 和: $k(x, x') = k_1(x, x') + k_2(x, x')$
 - 積: $k(x, x') = k_1(x, x') k_2(x, x')$

$\mathbf{z} = [x, y]^T$ の時:

- 和: $k(\mathbf{z}, \mathbf{z}') = k_1(x, x') + k_2(y, y')$
- 積: $k(\mathbf{z}, \mathbf{z}') = k_1(x, x') k_2(y, y')$

目次

- ガウス過程 (Gaussian Process; GP)
 - 序論
 - GPによる回帰
 - GPによる識別
- GP状態空間モデル
 - 概括
 - GP状態空間モデルによる音楽ムードの推定

線形モデルからGPへ(再び)

- 線形モデル

$$\mathbf{y} = \Phi \mathbf{w}, \quad \Phi = [\phi(x_1), \dots, \phi(x_N)]^T, \quad p(\mathbf{w}) = \mathbf{N}(\mathbf{w} | 0, \Sigma_{\mathbf{w}})$$



$$p(\mathbf{y} | \mathbf{x}) = \mathbf{N}(\mathbf{y} | 0, \mathbf{K})$$

$$\text{Gram matrix: } [\mathbf{K}]_{n,n'} = \phi(x_n)^T \phi(x_{n'}) = k(x_n, x_{n'})$$

GP: ノンパラメトリックモデル

GPによる回帰

- Prediction problem using a dataset $D = \{\mathbf{x}, \mathbf{y}\}$:

$$p(\mathbf{y}, y_* | \mathbf{x}, x_*) = \mathbf{N}(\mathbf{y}, y_* | \mathbf{0}_{N+1}, \mathbf{K}^+)$$

$\mathbf{K}^+ \equiv$

$\mathbf{K}_{\mathbf{x}, \mathbf{x}}$	$\mathbf{K}_{\mathbf{x}, x_*}$
$\mathbf{K}_{x_*, \mathbf{x}}$	\mathbf{K}_{x_*, x_*}

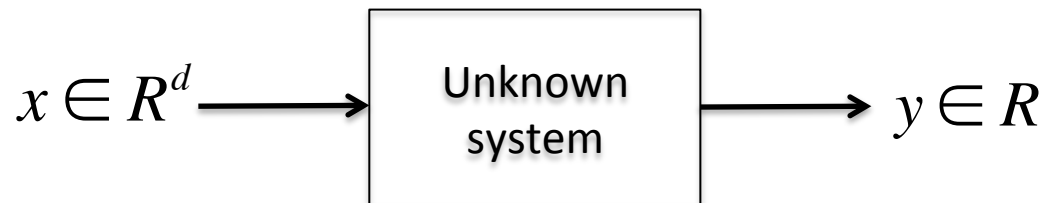


$$p(y_* | x_*, D) = \mathbf{N}(y_* | \mathbf{K}_{x_*, \mathbf{x}} \mathbf{K}_{\mathbf{x}, \mathbf{x}}^{-1} \mathbf{y}, \mathbf{K}_{x_*, x_*} - \mathbf{K}_{x_*, \mathbf{x}} \mathbf{K}_{\mathbf{x}, \mathbf{x}}^{-1} \mathbf{K}_{\mathbf{x}, x_*})$$

Predictive distribution

GPによる回帰

- 回帰問題: $y = f(x) + \varepsilon, \quad \varepsilon \sim \mathbf{N}(0, \sigma^2)$



- GP回帰:

$$\mathbf{f} = [f(x_1), f(x_2), \dots, f(x_n)]_{n=1, \dots, \infty}, \quad \mathbf{f} | \mathbf{x} \sim \mathbf{N}(\mathbf{m}, \mathbf{K})$$

$$f(x) \sim \mathcal{GP}(m(x), k(x, x'))$$

$$m(x) = E[f(x)]$$

$$k(x, x') = E[(f(x) - m(x))(f(x') - m(x')))]$$

GPによる回帰

- 平均と分散:

$$\mathbf{m} = [m(x_1), m(x_2), \dots, m(x_n)]$$

$$\mathbf{K} = \begin{bmatrix} k(x_1, x_1) & k(x_1, x_2) & \dots & k(x_1, x_n) \\ k(x_2, x_1) & k(x_2, x_2) & \dots & k(x_2, x_n) \\ \dots & \dots & \dots & \dots \\ k(x_n, x_1) & k(x_n, x_2) & \dots & k(x_n, x_n) \end{bmatrix}$$

$k(x, x')$ - any valid kernel function.

GPによる回帰

$$p(y_* | x_*, \mathbf{y}, \mathbf{X}) = \int \underbrace{p(y_* | f_*)}_{\text{Gaussian}} \underbrace{p(f_* | x_*, \mathbf{y}, \mathbf{X})}_{\text{Gaussian}} df_*$$

$$\underbrace{p(f_* | x_*, \mathbf{y}, \mathbf{X})}_{\text{Gaussian}} = \int \underbrace{p(f_* | \mathbf{f}, x_*, \mathbf{X})}_{\text{Gaussian}} \underbrace{p(\mathbf{f} | \mathbf{y}, \mathbf{X})}_{\text{Gaussian}} d\mathbf{f}$$

$$p(y_* | x_*, \mathbf{y}, \mathbf{X}) = \mathbf{N}(\bar{f}_*, \text{var}(f_*))$$

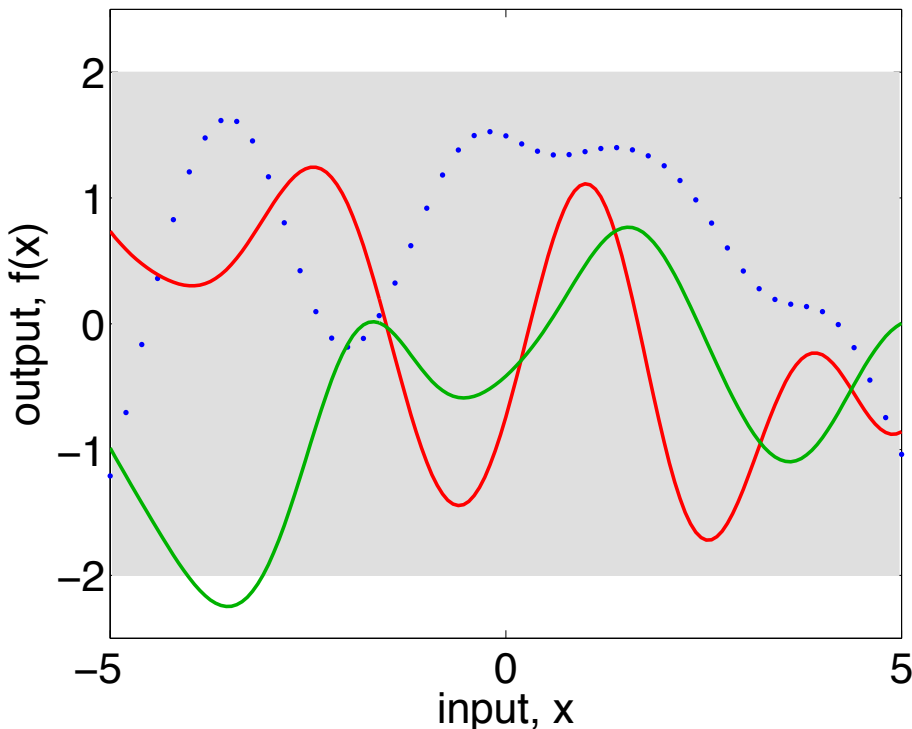
GPによる回帰

- 回帰問題: $y = f(x) + \varepsilon, \quad \varepsilon \sim \mathbf{N}(0, \sigma^2)$
- 予測: $p(y_* | x_*, \mathbf{y}, \mathbf{x}) = \mathbf{N}(\bar{f}_*, \text{var}(f_*) + \sigma^2 \mathbf{I})$

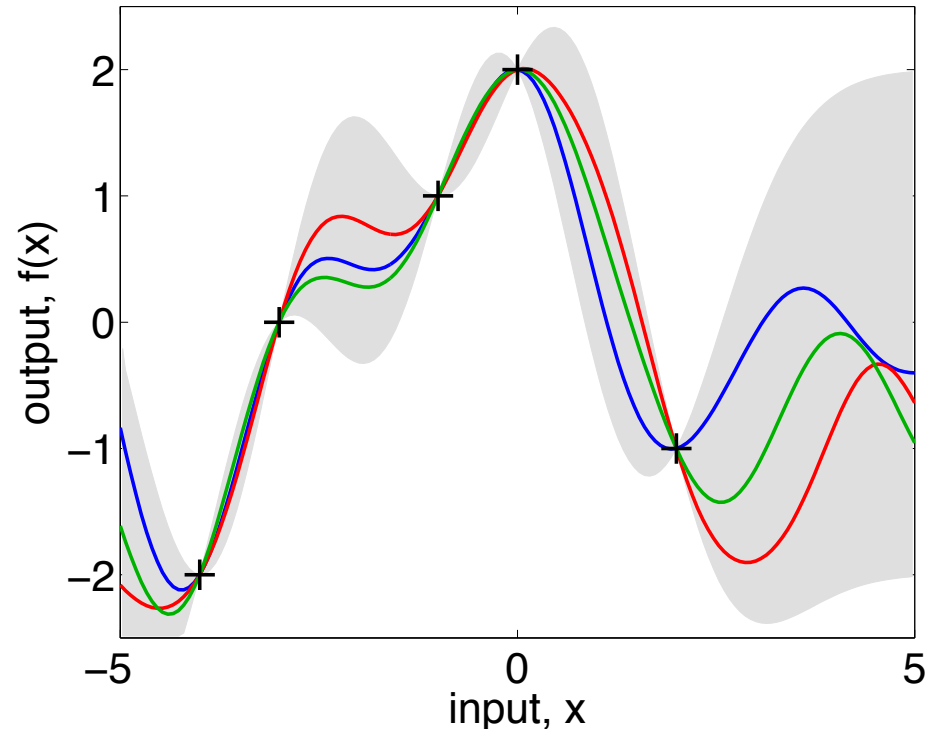
$$\bar{f}_* = \mathbf{K}_{x_*, \mathbf{x}} (\mathbf{K}_{\mathbf{x}, \mathbf{x}} + \sigma^2 \mathbf{I})^{-1} \mathbf{y}$$

$$\text{var}(f_*) = \mathbf{K}_{x_*, x_*} - \mathbf{K}_{x_*, \mathbf{x}} (\mathbf{K}_{\mathbf{x}, \mathbf{x}} + \sigma^2 \mathbf{I})^{-1} \mathbf{K}_{\mathbf{x}, x_*}$$

Prior



Posterior



GPによる回帰

- ハイパーパラメータの学習:

$$\Theta = \{\sigma, l\}$$

$$\max_{\Theta} p(\mathbf{y} | \mathbf{x}, \Theta) = \max_{\Theta} \int p(\mathbf{y} | \mathbf{f}) p(\mathbf{f} | \mathbf{x}, \Theta) d\mathbf{f}$$

ML type -II estimation

目次

- ガウス過程 (Gaussian Process; GP)
 - 序論
 - GPによる回帰
 - GPによる識別
- GP状態空間モデル
 - 概括
 - GP状態空間モデルによる音楽ムードの推定

GPによる識別

- 入力 x のクラス c を推定する問題:

$$p(c | x) = \int p(c | y, x) p(y | x) dy = \int p(c | y) p(y | x) dy$$

- 学習データ $\mathbf{X} = \{x^1, \dots, x^N\}$ とそのクラス $\mathbf{C} = \{c^1, \dots, c^N\}$ が与えられた時、新しい入力 x^* のクラスを推定する:

$$p(c^* | x^*, \mathbf{C}, \mathbf{X}) = \int p(c^* | y^*) p(y^* | \mathbf{X}, \mathbf{C}) dy^*$$

$$p(y^* | \mathbf{X}, \mathbf{C}) \propto p(y^*, \mathbf{C} | \mathbf{X}) = \int p(y^*, \mathbf{Y}, \mathbf{C} | \mathbf{X}, x^*) d\mathbf{Y}$$

$$= \int p(\mathbf{C} | \mathbf{Y}) p(\mathbf{Y}, y^* | \mathbf{X}, x^*) d\mathbf{Y}$$

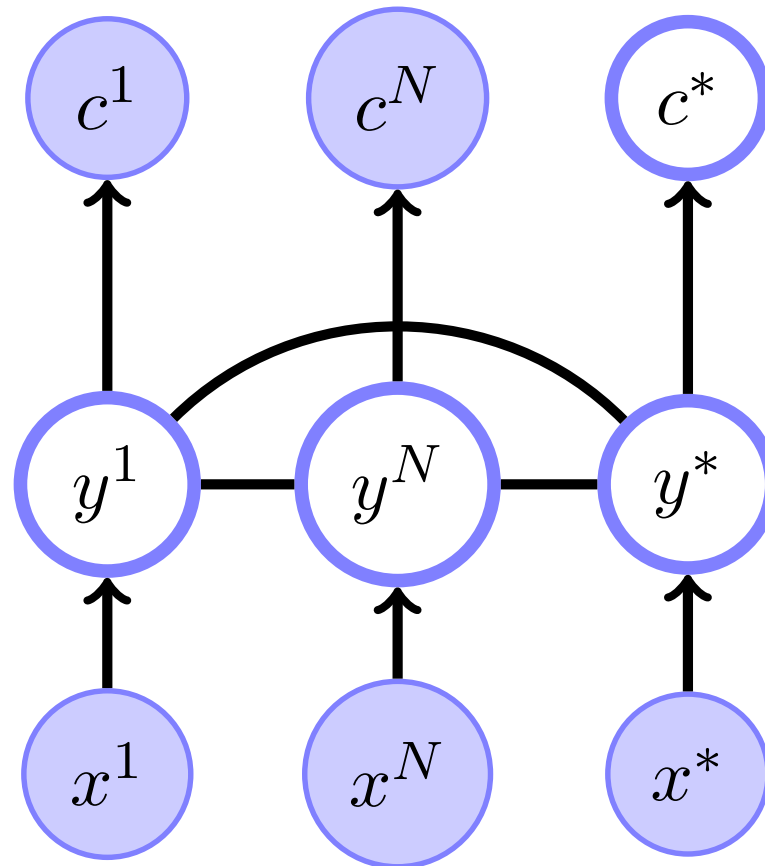
クラス写像

$$= \int \left\{ \prod_{n=1}^N p(c^n | y^n) \right\} p(y^1, \dots, y^N, y^* | x^1, \dots, x^N, x^*) dy^1, \dots, dy^N$$

GP

GPによる識別

- グラフ表現:



GPによる識別: 2クラス分類 $c \in \{0, 1\}$

例) クラス写像としてシグモイド関数を利用:

$$\sigma(x) = \frac{1}{1 + \exp(-x)}$$

$$p(c | y) = \sigma((2c - 1)y)$$

【問題】非線形のクラス写像の場合、 $p(y^* | \mathbf{X}, \mathbf{C})$ の積分計算

$$p(y^* | \mathbf{X}, \mathbf{C}) \propto \int \left\{ \prod_{n=1}^N p(c^n | y^n) \right\} p(\mathbf{Y}, y^* | \mathbf{X}, x^*) d\mathbf{Y}$$

GPによる識別: 2クラス分類 $c \in \{0, 1\}$

- ラプラス近似法 (分布のGaussian近似法) の利用:

$$p(y^*, \mathbf{Y} | x^*, \mathbf{X}, \mathbf{C}) \propto p(y^*, \mathbf{Y}, \mathbf{C} | x^*, \mathbf{X})$$

$$\propto p(y^* | \mathbf{Y}, x^*, \mathbf{X}) p(\mathbf{Y} | \mathbf{C}, \mathbf{X})$$

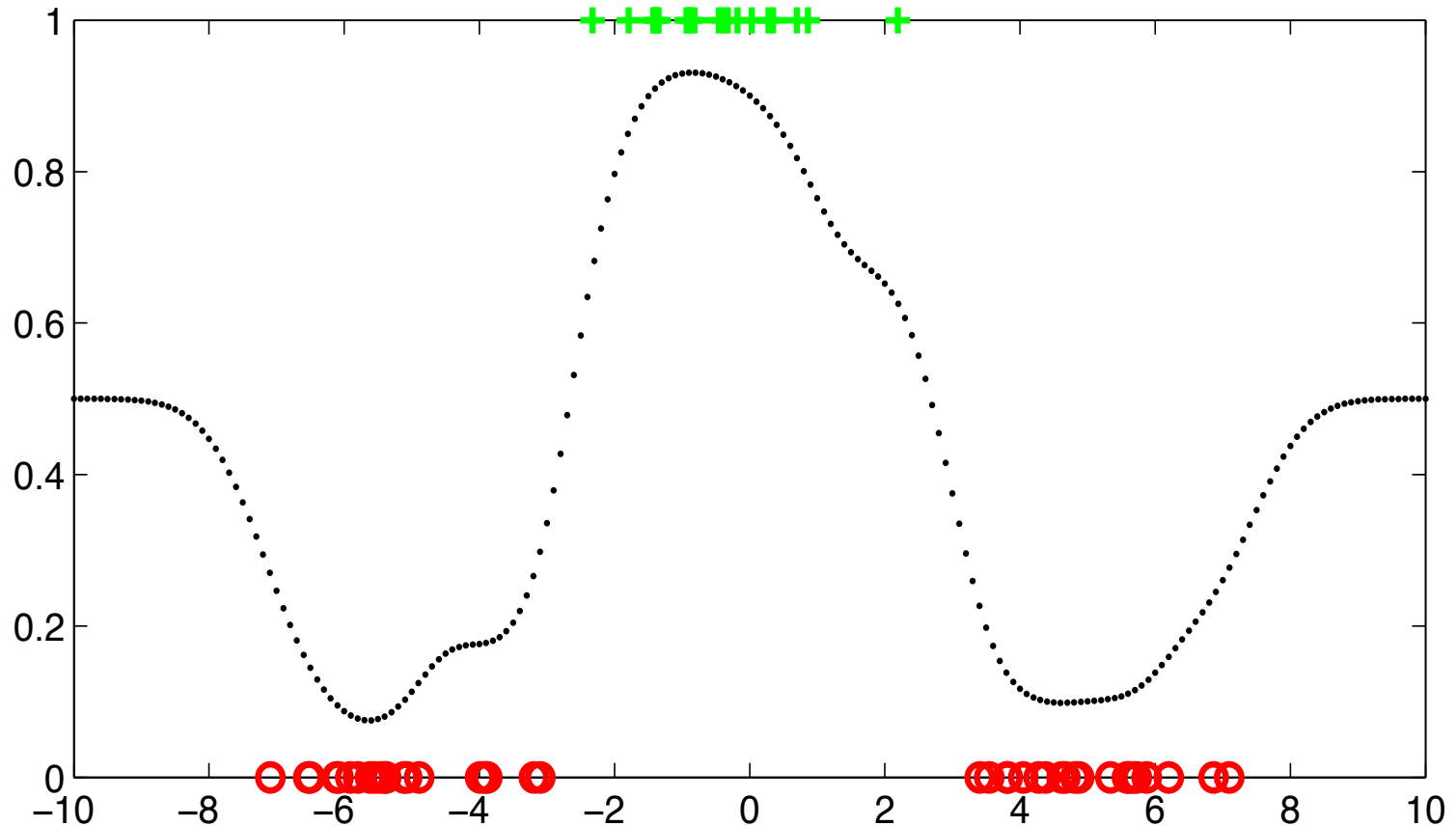
クラス情報を含まない Gaussianで近似

$$p(y^*, \mathbf{Y} | x^*, \mathbf{X}, \mathbf{C}) \approx p(y^* | \mathbf{Y}, x^*, \mathbf{X}) q(\mathbf{Y} | \mathbf{C}, \mathbf{X})$$

Joint Gaussian

$$p(c^* = 1 | x^*, \mathbf{X}, \mathbf{C}) \approx \left\langle \sigma(y^*) \right\rangle_{N(y^* | \langle y^* \rangle, \text{var}(y^*))}$$

GPによる識別: 2クラス分類 $c \in \{0,1\}$



GPによる識別：多クラス分類

- ソフトマックス関数の利用：

$$p(c = m | y) = \frac{\exp(y_m)}{\sum_{m'} \exp(y_{m'})}$$

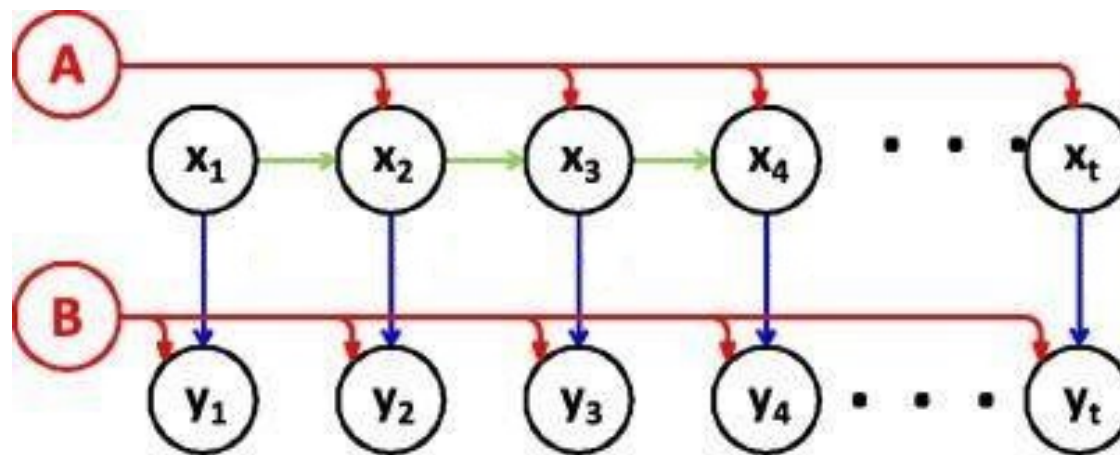
$$\sum_m p(c = m) = 1$$

目次

- ガウス過程 (Gaussian Process; GP)
 - 序論
 - GPによる回帰
 - GPによる識別
- GP状態空間モデル
 - 概括
 - GP状態空間モデルによる音楽ムードの推定

状態空間モデル

- 時系列を解析するモデル: 状態 x_t と観測 y_t を表す式



$$\begin{cases} x_t = f(x_{t-1}, A) + v_t, & v \sim \mathcal{N}(0, R) \\ y_t = g(x_t, B) + v_t, & v \sim \mathcal{N}(0, Q) \end{cases}$$

カルマンフィルター

- 線形状態空間モデル

$$\begin{cases} x_t = Fx_{t-1} + v_t \\ y_t = Gx_t + v_t \end{cases}$$

- 長所:
 - 解析的な推定法が確立されている
 - 高速に計算できる
- 短所:
 - 線形のモデル

GP状態空間モデル

- 状態と観測の式:

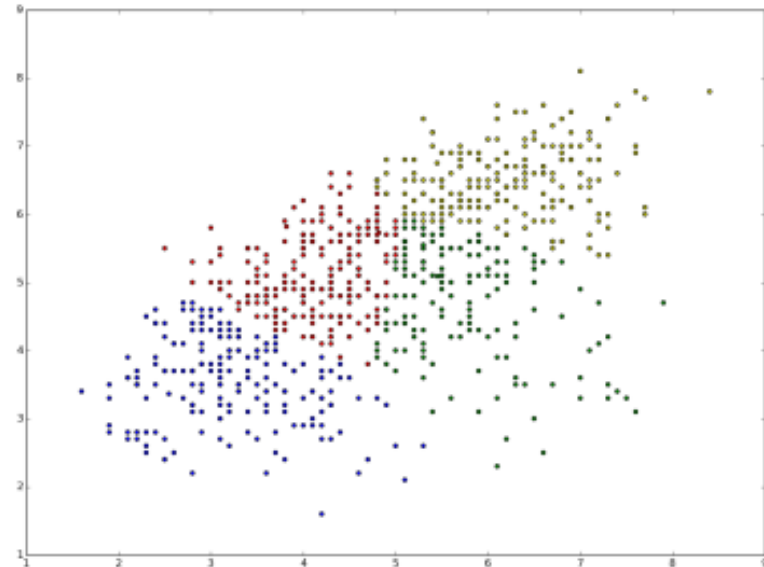
$$\begin{cases} x_t = f(x_{t-1}) + v_t, & f(x) \sim \mathcal{GP}(0, K_f) \\ y_t = g(x_t) + v_t, & g(x) \sim \mathcal{GP}(0, K_g) \end{cases}$$

- 長所:
 - 非線形、ノンパラメトリックのモデル
 - 表現力が高い
- 短所:
 - 推定法がまだ確立されていない
 - 計算コストが大きい

実験

MediaEval2014

- 使用データ
 - 学習セット - 600 clips.
 - テストセット - 144 clips.
- カルマンフィルター、GP状態空間モデル
 - A-V空間を4つにクラスタリング
 - 各クラスごとに状態空間モデルを推定
 - 最尤基準によるモデル選択



実験

MediaEval2014

- 音楽特徴量
 - mfcc – メル周波数ケプストラム係数
 - baseline – MediaEval2014における標準の特徴量
(スペクトルの変化、ハーモニックの変化、音量、音色、ゼロ交差率を表す5つの特徴量)

実験結果 : RMSE

特徴量	カルマンフィルター	GP状態空間モデル
	RMSE	RMSE
AROUSAL – MULTIPLE MODELS		
mfcc	0.0852	0.1089
baseline	0.3824	0.2073
VALENCE – MULTIPLE MODELS		
mfcc	0.1590	0.1096
baseline	0.2325	0.2267

ご質問をどうぞ！