Expectation Propagation for t-Exponential Family Using q-Algebra

Futoshi Futami
University of Tokyo
RIKEN

Issei Sato
University of Tokyo
RIKEN

Masashi Sugiyama
RIKEN
University of Tokyo

2017, Nov.8 IBIS

Gaussian distribution

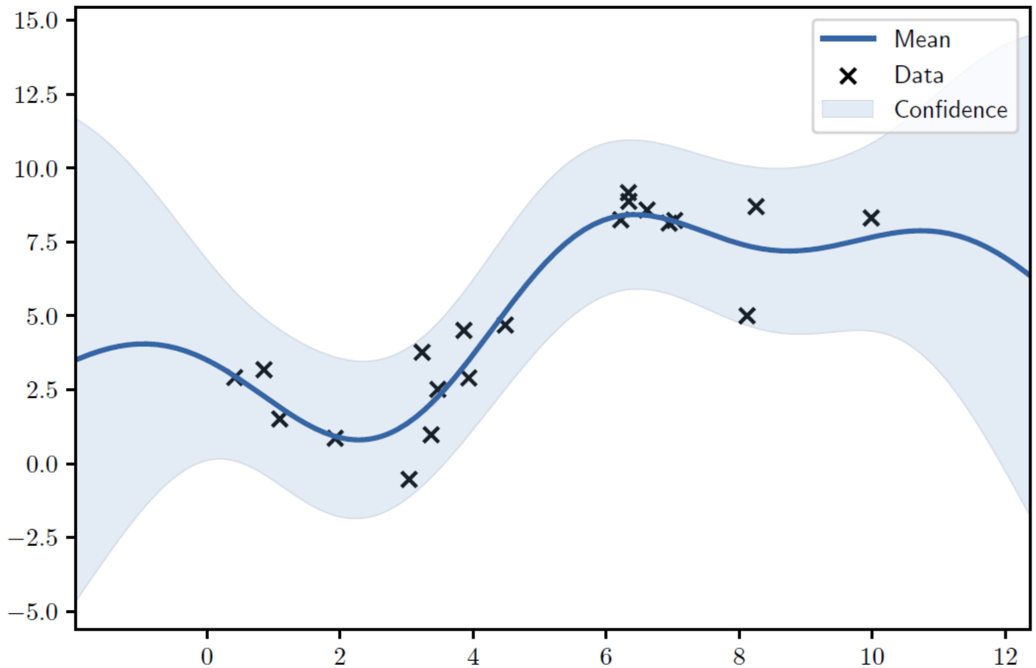
$$\mathcal{N}(\mu, \Sigma) = \frac{1}{\sqrt{(2\pi)^d |\Sigma|}} \exp\left\{-\frac{1}{2}(x-\mu)^\top \Sigma^{-1}(x-\mu)\right\}$$

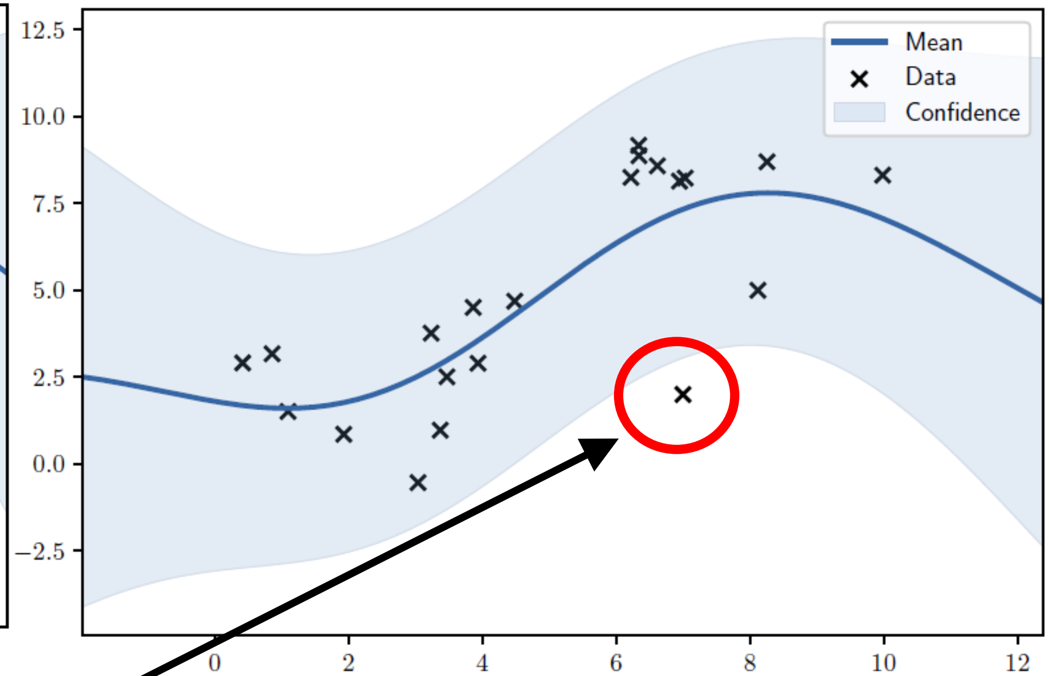
- **Plays central role in ML**
- **E.g. Gaussian process**

Sensitive to outliers

likelihood	**Prior**
$y_i = f(x_i) + \epsilon_i$	$p(f\|X) = \mathcal{N}(f\|\mu, K)$
$p(y_i\|f_i) = \mathcal{N}(y_i\|f_i, \epsilon_i)$	$[K]_{ij} = k(x_i, x_j)$

Example of Gaussian process (GP) regression



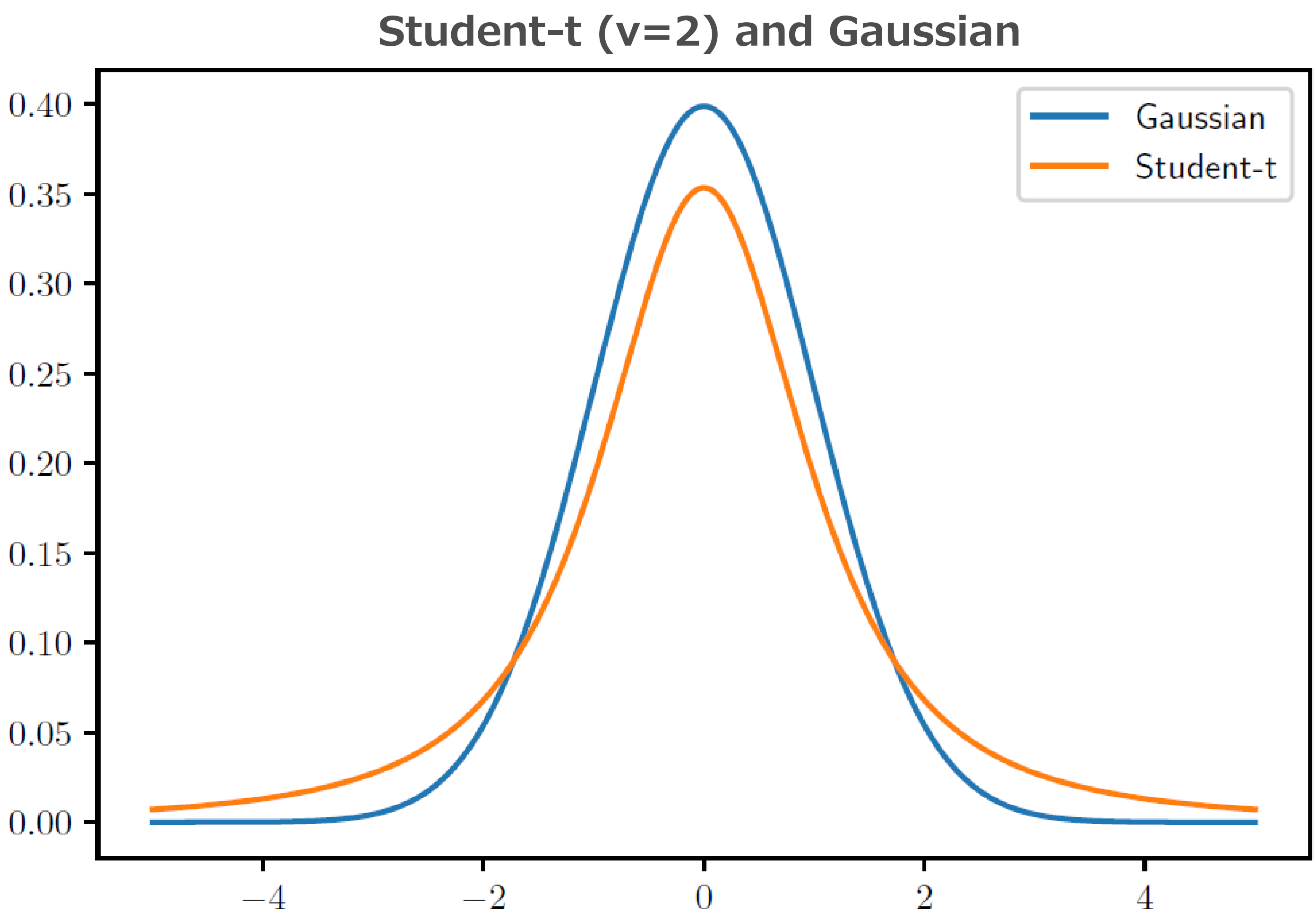


If an outlier exists…

GP is sensitive to outliers…

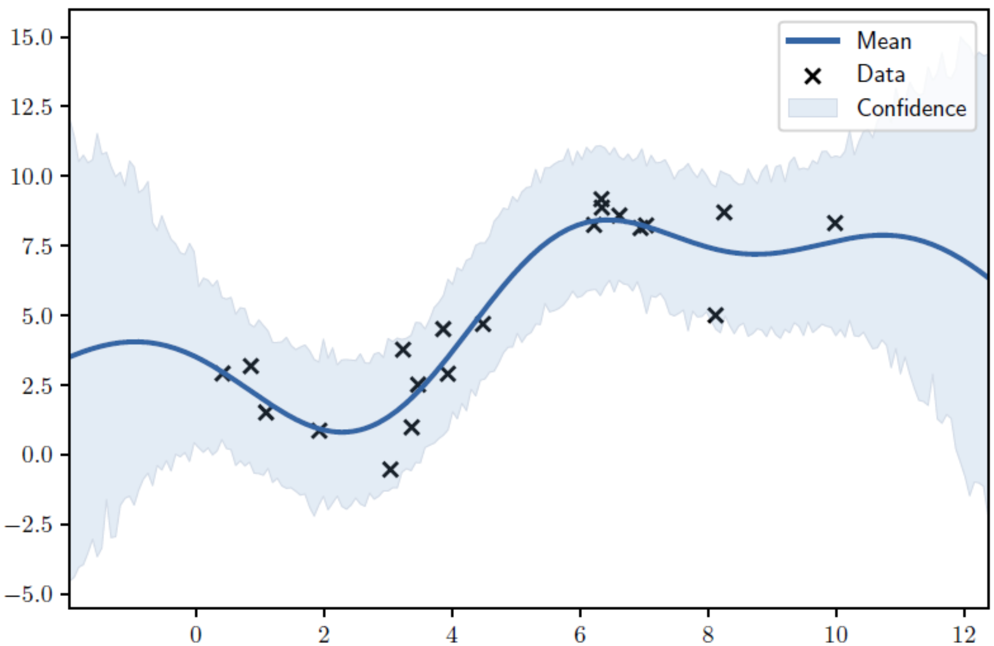
Student-t distribution

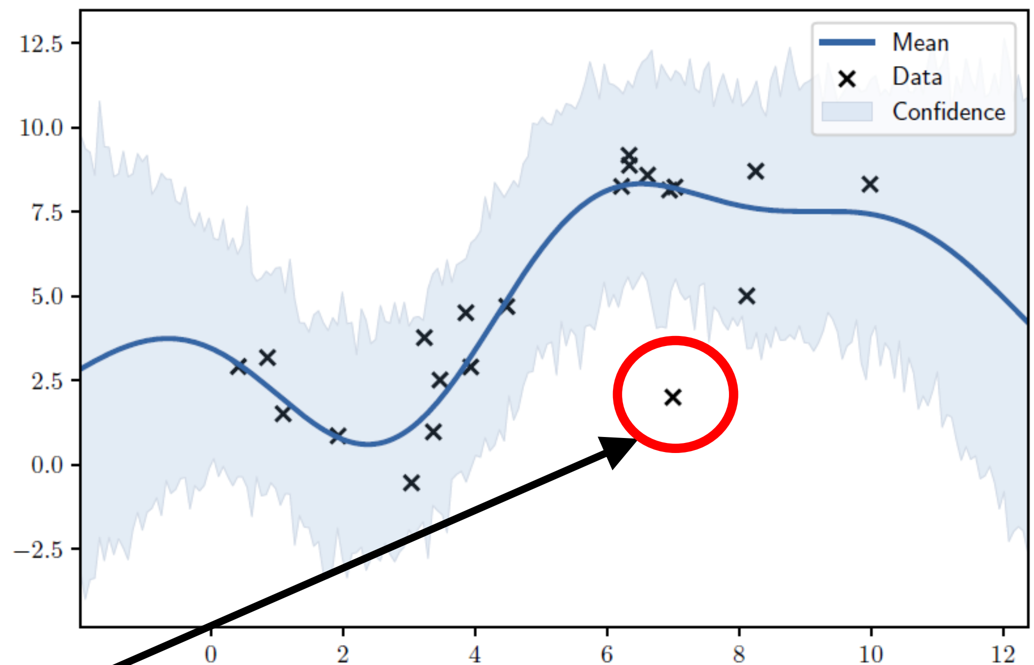
$$\mathrm{St}(x; v, \mu, \Sigma) = \frac{\Gamma((v+k)/2)}{(\pi v)^{k/2}\Gamma(v/2)|\Sigma|^{1/2}}\left(1 + (x-\mu)^\top (v\Sigma)^{-1}(x-\mu)\right)^{-\frac{v+k}{2}}$$



GP example（Student-t likelihood）

Replace Gaussian likelihood with Student-t





Student-t likelihood is less affected by outliers

Importance of Student-t as an alternative to Gaussian

The problem of dealing with Student-t

Student-t is difficult to handle...

- **Not a member of the Exponential family.**
- **We cannot use many useful relations which make Gaussian easy to use.**
- **Need to approximate Student–t further by another distribution which is easy to handle (variational inference, EP, Laplace)**

How to use Student-t directly as an alternative to Gaussian?

t-exponential family

[N. Ding and S. V. N. Vishwanathan, 2010,2011]

$$p(x;\theta) = \exp(\langle \Phi(x), \theta \rangle - g(\theta))$$

θ : natural parameter

$\Phi(x)$: sufficient statistics

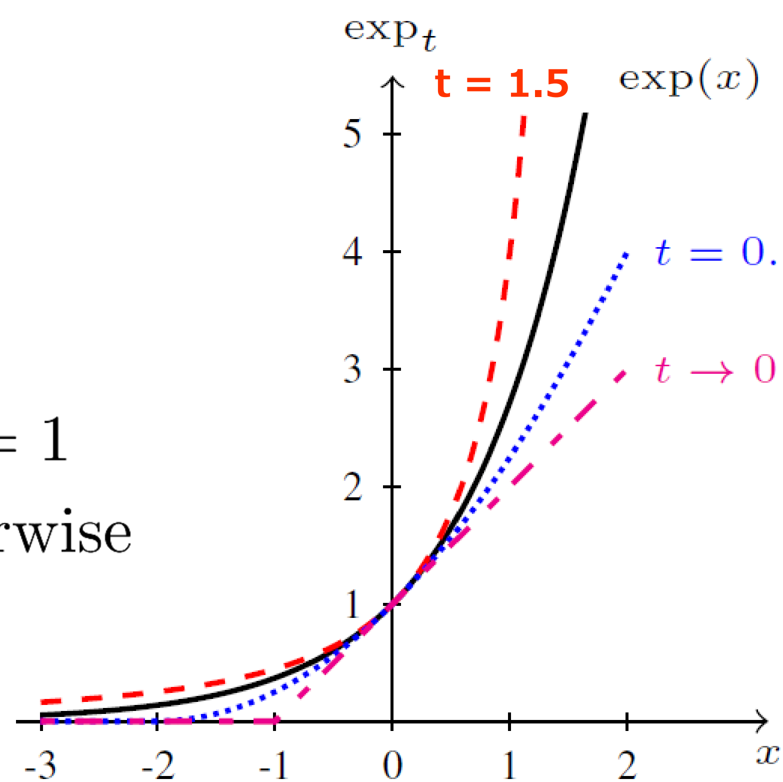
$g_t(\theta)$: log partition function

$\boxed{\textbf{Extend Exponential Family}}$

$$p(x;\theta) = \exp_t(\langle \Phi(x), \theta \rangle - g_t(\theta))$$

- **Deformed exponential**

$$\exp_t(x) = \begin{cases} \exp(x) & \text{if } t = 1 \\ [1 + (1-t)x]^{\frac{1}{1-t}} & \text{otherwise} \end{cases}$$



Approximation for t-exponential family

[N. Ding and S. V. N. Vishwanathan, 2010,2011]

exponential family	t-exponential family
Assumed Density Filtering	**Assumed Density Filtering**
Online algorithm	Online algorithm

- **The performance heavily depends on the order in which data is processed.**
- **Sometimes disastrous results are obtained.**

Expectation Propagation

[T. Minka, 2001]

exponential family

Assumed Density Filtering

Data: permutation dependent

Calculation through natural parameter

Expectation Propagation

Data: permutation independent

Property of exponential family

- **A product of same exponential families yields the same un-normalized exponential family**

$$e^{\langle \Phi(x), \theta_1 \rangle - g(\theta_1)} e^{\langle \Phi(x), \theta_2 \rangle - g(\theta_2)}$$

$$= e^{\langle \Phi(x), (\theta_1 + \theta_2) \rangle - \tilde{g}(\theta_1, \theta_2)}$$

- **Calculation can be performed through natural parameter or expected sufficient statistics**

Approximation for t-exponential family

[T. Minka, 2001]

exponential family

Assumed Density Filtering

Data: permutation dependent

Calculation through natural parameter

Expectation Propagation

Data: permutation independent

[N. Ding and S. V. N. Vishwanathan, 2010,2011]

t-exponential family

Assumed Density Filtering

Data: permutation dependent

Problem of t-Exponential family

$$\exp_t(x)\exp_t(y) \neq \exp_t(x+y)$$

A product of t-exponential families does not yield the same un-normalized t-exponential family

Calculation cannot be performed through natural parameter or expected sufficient statistics

Expectation Propagation

[T. Minka, 2001]

exponential family

Assumed Density Filtering

Data: permutation dependent

Calculation through natural parameter

Expectation Propagation

Data: permutation independent

[N. Ding and S. V. N. Vishwanathan, 2010,2011]

t-exponential family

Assumed Density Filtering

Data: permutation dependent

~~**Calculation through natural parameter**~~

Expectation Propagation

Data: permutation independent

Our Contribution

t-exponential family

Assumed Density Filtering

Data: permutation dependent

① Calculation through natural parameter

②

Expectation Propagation

Data: permutation independent

By using q-algebra

①

Enabled natural parameter based calculation in t-exponential family

②

Extend ADF to EP in t-exponential family

q-Algebra

[H. Suyari and M. Tsukada, 2005] [L. Nivanen, A. Le Mehaute, and Q. A. Wang, 2003]

We can treat the pseudo additivity efficiently by using q-algebra

$$\exp_t(x)\exp_t(y) = \exp_t(x + y + (1-t)xy)$$

Definition 1 (q-product). *Operation* $\otimes_q$ *called* q-product *is defined as*

$$x \otimes_q y := \begin{cases} [x^{1-q} + y^{1-q} - 1]^{\frac{1}{1-q}} & \text{if } x > 0, y > 0, \\ & \quad x^{1-q} + y^{1-q} - 1 > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Definition 2 (q-division). *Operation* $\oslash_q$ *called* q-division *is defined as*

$$x \oslash_q y := \begin{cases} [x^{1-q} - y^{1-q} + 1]^{\frac{1}{1-q}} & \text{if } x > 0, y > 0, \\ & \quad x^{1-q} - y^{1-q} - 1 > 0, \\ 0 & \text{otherwise.} \end{cases}$$

Numerical Experiment

Comparison of binary classification by using Student-t Process and Gaussian Process

	Student-t Process classifier	Gaussian Process classifier
prior	$p(f\|X) = \mathrm{St}(f\|v, \mu, K)$	$p(f\|X) = \mathcal{N}(f\|\mu, K)$
likelihood	$p(y_i\|f_i) = \epsilon + (1-2\epsilon)\Theta(y_i f_i)$	

[H. C. Kim and Z. Ghahramani 2008]

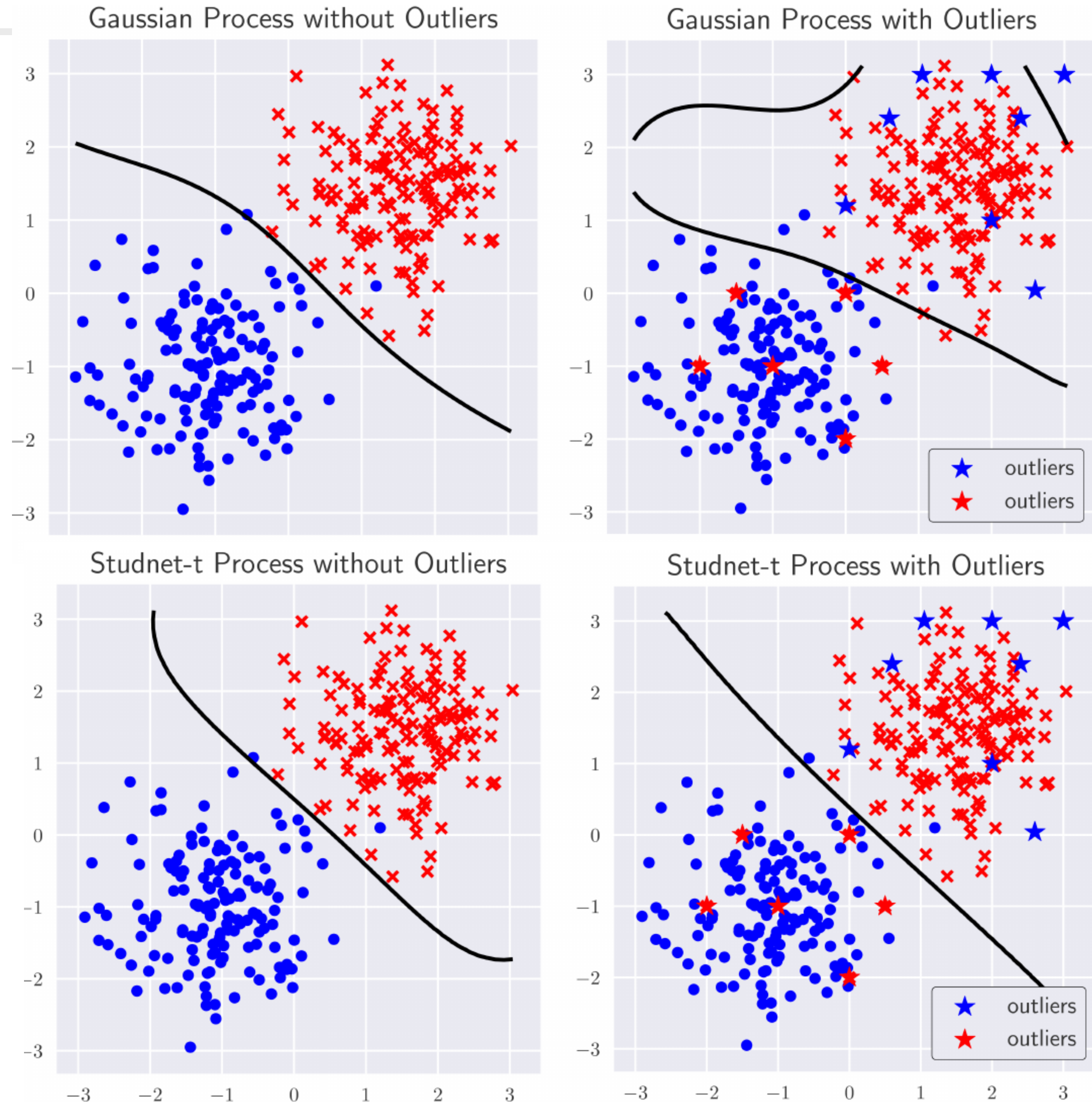
① **Toy datasets**

How outliers affect the decision boundary

② **UCI datasets**

Compare the performance

Experiment① toy experiment



Student–t is less affected by outliers than Gaussian

Experiment② UCI benchmark data

Table 1: Classification Error Rates(%)

Dataset	Outliers	GPC	STC
Pima	0	34.0(3.0)	**32.3(2.6)**
	5%	34.9(3.1)	**32.9(3.1)**
	10%	36.2(3.3)	**34.4(3.5)**
Ionosphere	0	9.6(1.7)	**7.5(2.0)**
	5%	9.9(2.8)	**9.6(3.2)**
	10%	13.0(5.2)	**11.9(5.4)**
Thyroid	0	**4.3(1.3)**	4.4(1.3)
	5%	**4.8(1.8)**	5.5(2.3)
	10%	**5.4(1.4)**	7.2(3.4)
Sonar	0	15.4(3.6)	**15.0(3.2)**
	5%	18.3(4.4)	**17.5(3.3)**
	10%	**19.4(3.8)**	**19.4(3.1)**

Better performance compared to Gaussian process on many datasets

Summary

t-exponential family

Assumed Density Filtering

Data: permutation dependent

① Calculation through natural parameter

②

Expectation Propagation

Data: permutation independent

By using q-algebra

①
Enabled natural parameter based calculation in t-exponential family

②
Extend ADF to EP in t-exponential family