Gaussian Processes for Regression: Models, Algorithms, and Applications

Tamara Broderick
Associate Professor
MIT

http://www.tamarabroderick.com/tutorials.html

Why Gaussian processes (GPs)?

Why Gaussian processes (GPs)?

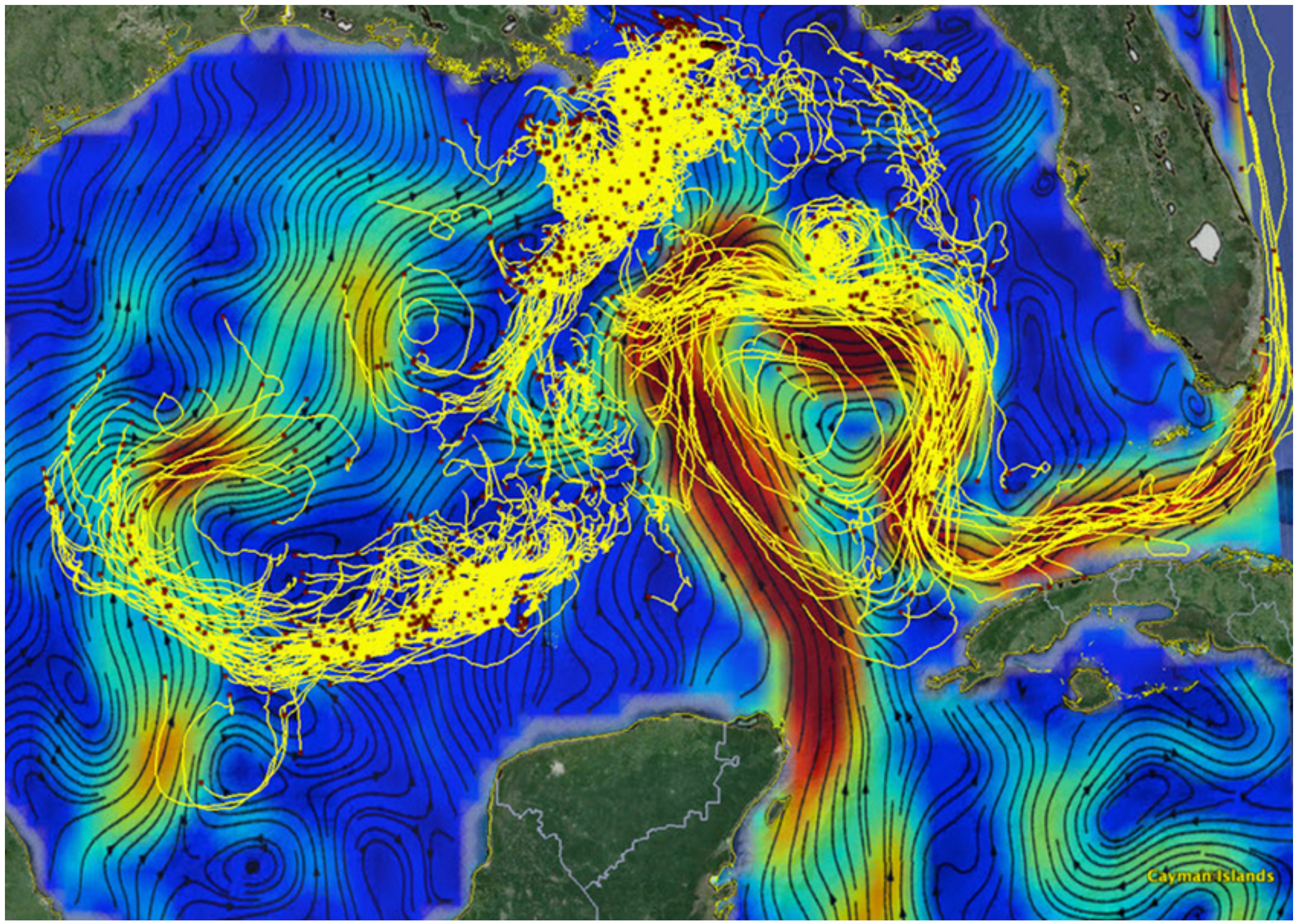
- Often want to estimate/“predict” some continuous outcome as a function of certain inputs (*regression*)

Why Gaussian processes (GPs)?

- Often want to estimate/"predict" some continuous outcome as a function of certain inputs (*regression*)
- GPs are good at certain types of regression problems

Why Gaussian processes (GPs)?

- Often want to estimate/“predict” some continuous outcome as a function of certain inputs (*regression*)
- GPs are good at certain types of regression problems

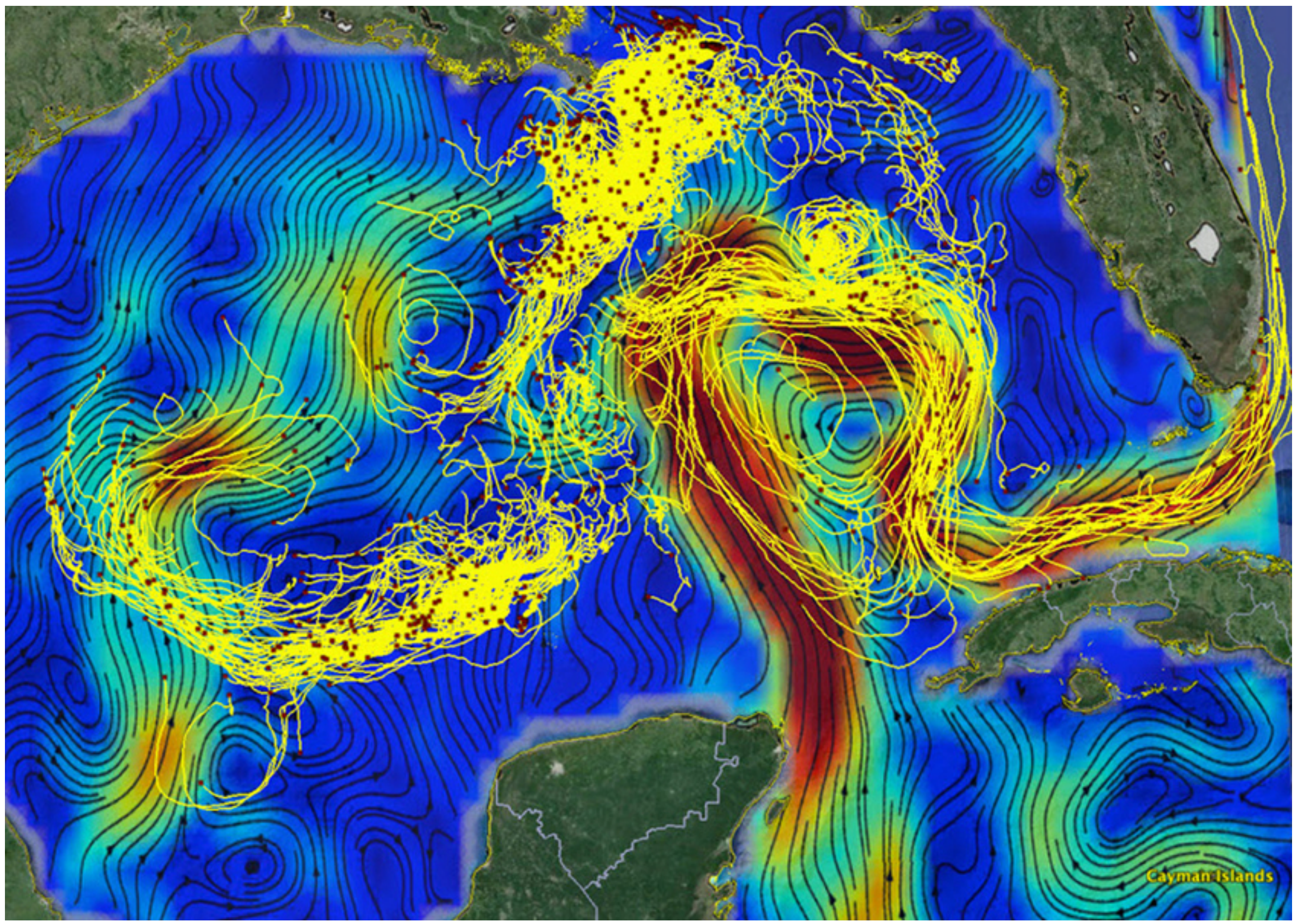


Example:

[Ryan, Özgökmen 2023; Zewe 2023; Gonçalves et al 2019; Lodise et al 2020; Berlinghieri et al 2023]

Why Gaussian processes (GPs)?

- Often want to estimate/“predict” some continuous outcome as a function of certain inputs (*regression*)
- GPs are good at certain types of regression problems



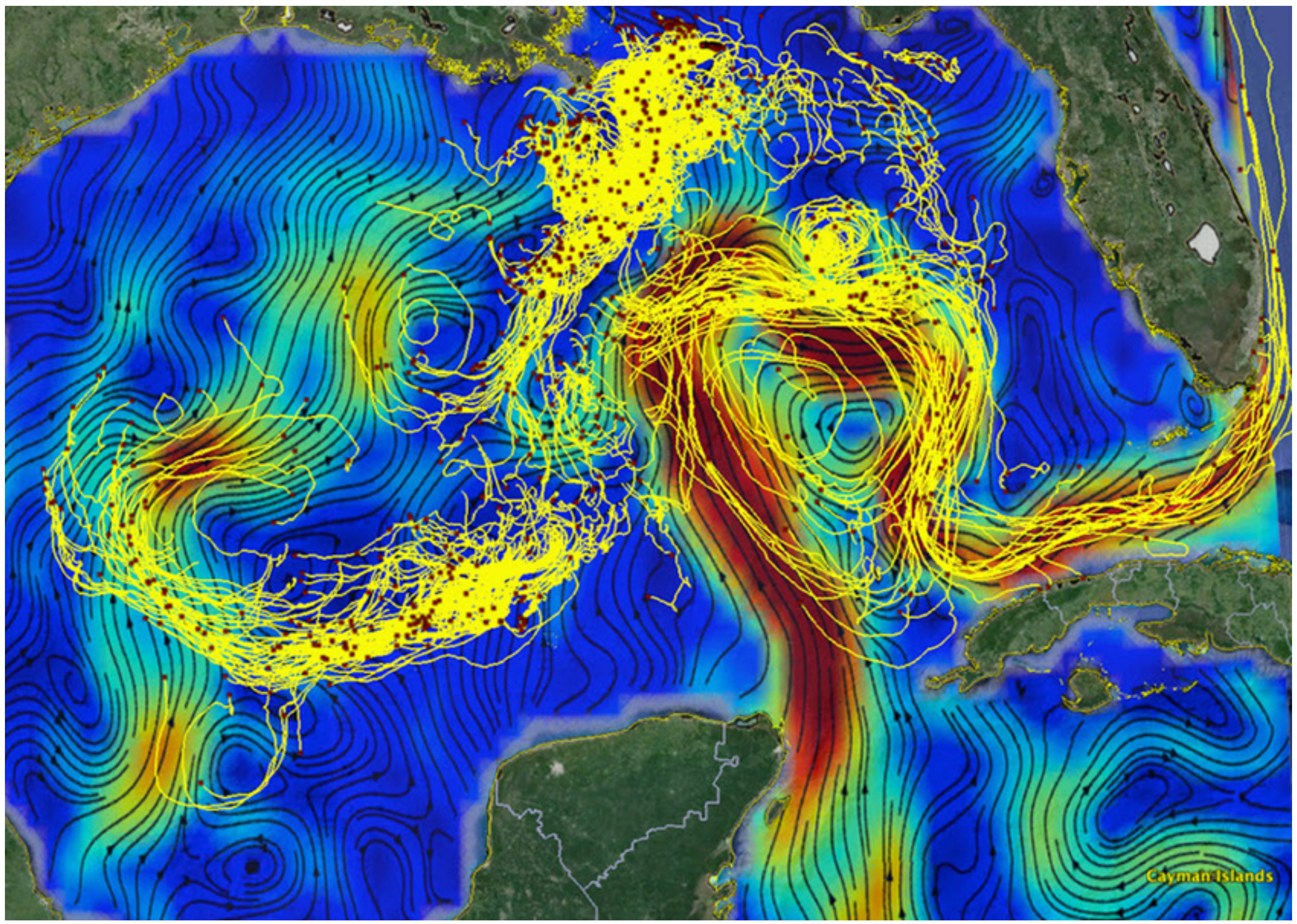
Example:

- The ocean current (velocity vector field) varies by space & time

[Ryan, Özgökmen 2023; Zewe 2023; Gonçalves et al 2019; Lodise et al 2020; Berlinghieri et al 2023]

Why Gaussian processes (GPs)?

- Often want to estimate/"predict" some continuous outcome as a function of certain inputs (*regression*)
- GPs are good at certain types of regression problems



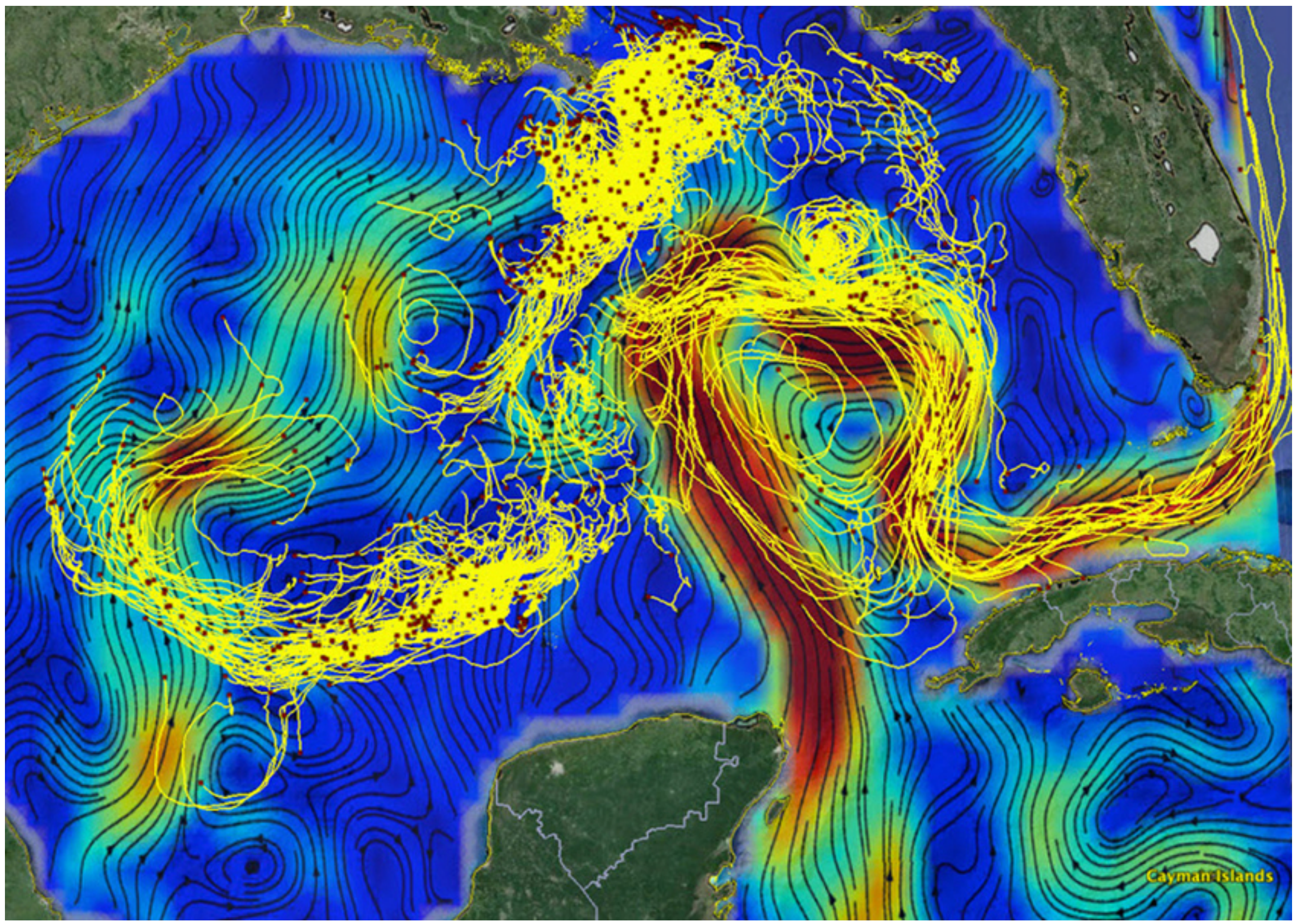
Example:

- The ocean current (velocity vector field) varies by space & time
- Scientists get sparse observations of the current from buoys

[Ryan, Özgökmen 2023; Zewe 2023; Gonçalves et al 2019; Lodise et al 2020; Berlinghieri et al 2023]

Why Gaussian processes (GPs)?

- Often want to estimate/“predict” some continuous outcome as a function of certain inputs (*regression*)
- GPs are good at certain types of regression problems



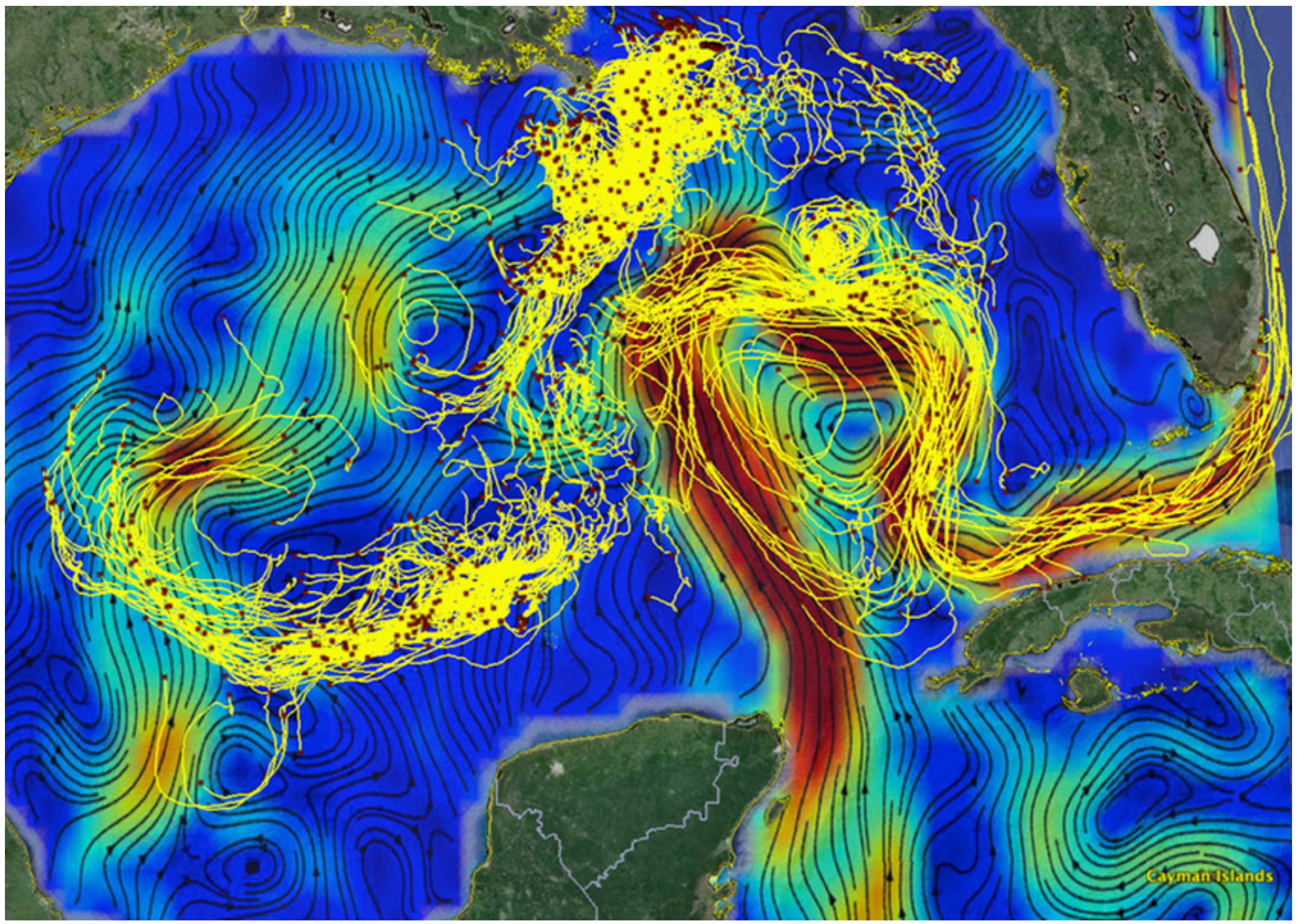
Example:

- The ocean current (velocity vector field) varies by space & time
- Scientists get sparse observations of the current from buoys
- Goal: estimate the current

[Ryan, Özgökmen 2023; Zewe 2023; Gonçalves et al 2019; Lodise et al 2020; Berlinghieri et al 2023]

Why Gaussian processes (GPs)?

- Often want to estimate/“predict” some continuous outcome as a function of certain inputs (*regression*)
- GPs are good at certain types of regression problems



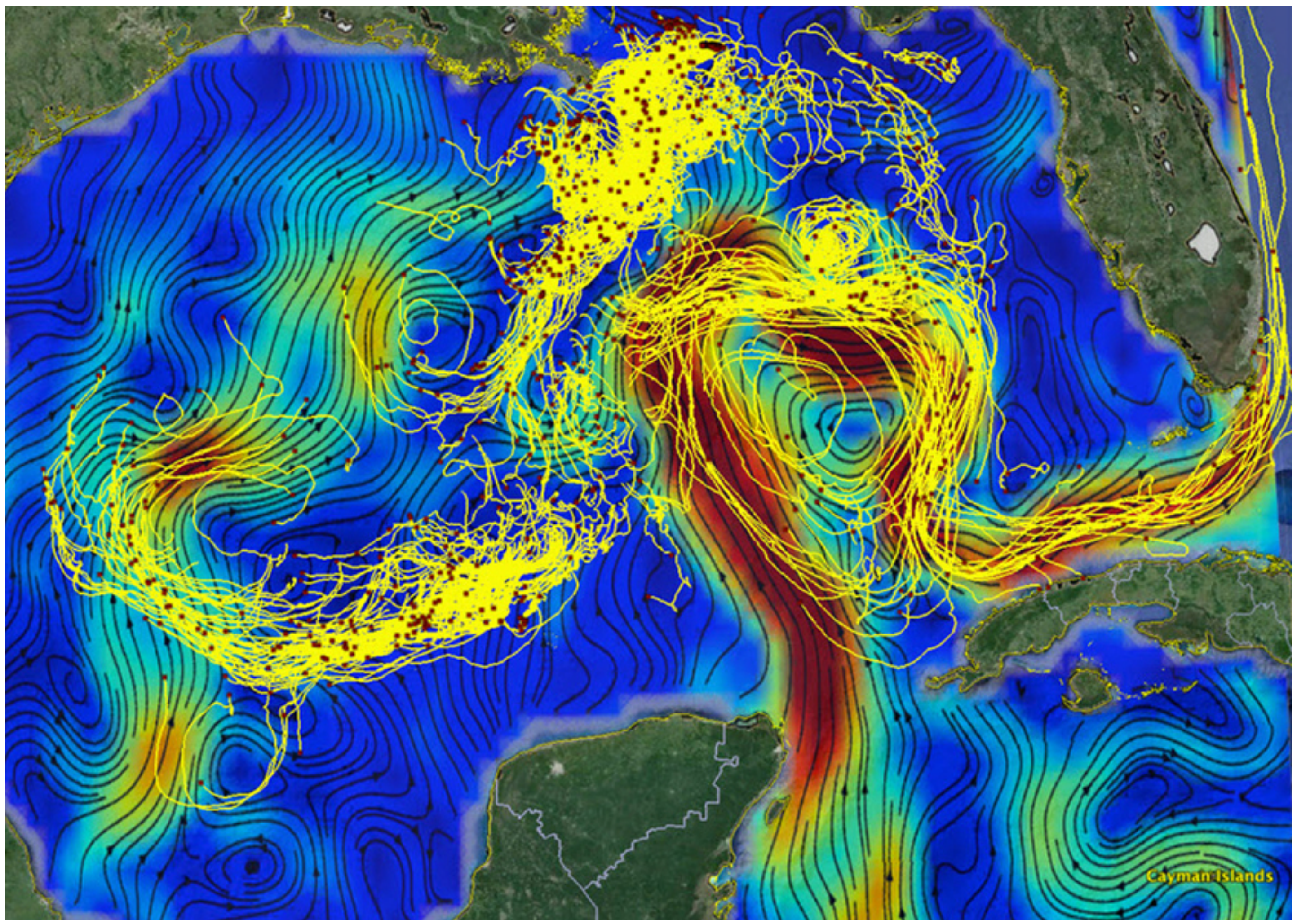
Example:

- The ocean current (velocity vector field) varies by space & time
- Scientists get sparse observations of the current from buoys
- Goal: estimate the current “predict”

[Ryan, Özgökmen 2023; Zewe 2023; Gonçalves et al 2019; Lodise et al 2020; Berlinghieri et al 2023]

Why Gaussian processes (GPs)?

- Often want to estimate/“predict” some continuous outcome as a function of certain inputs (*regression*)
- GPs are good at certain types of regression problems



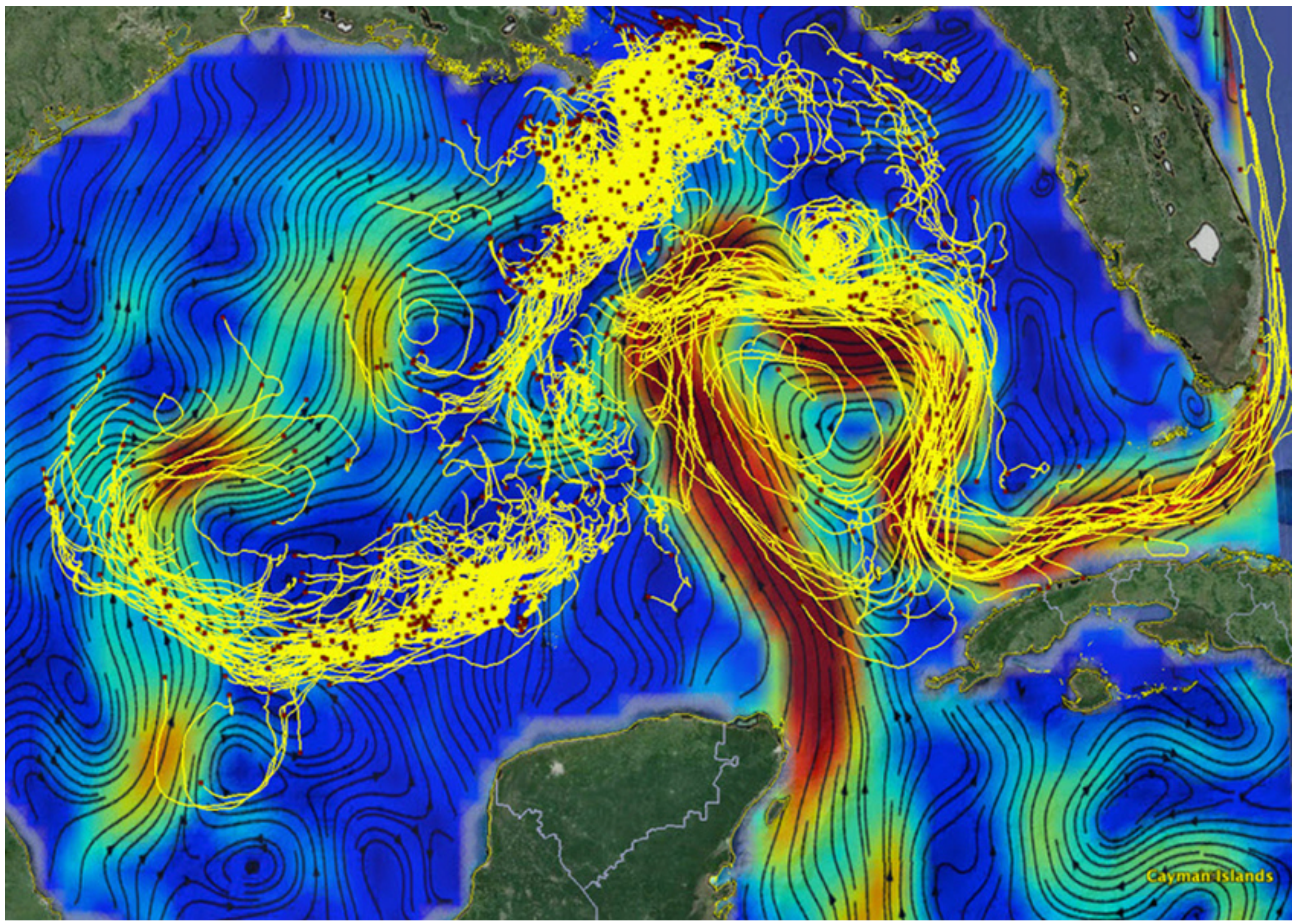
Example:

- The ocean current (velocity vector field) varies by space & time
- Scientists get sparse observations of the current from buoys
- Goal: estimate the current “predict”

[Ryan, Özgökmen 2023; Zewe 2023; Gonçalves et al 2019; Lodise et al 2020; Berlinghieri et al 2023]

Why Gaussian processes (GPs)?

- Often want to estimate/“predict” some continuous outcome as a function of certain inputs (*regression*)
- GPs are good at certain types of regression problems



Example:

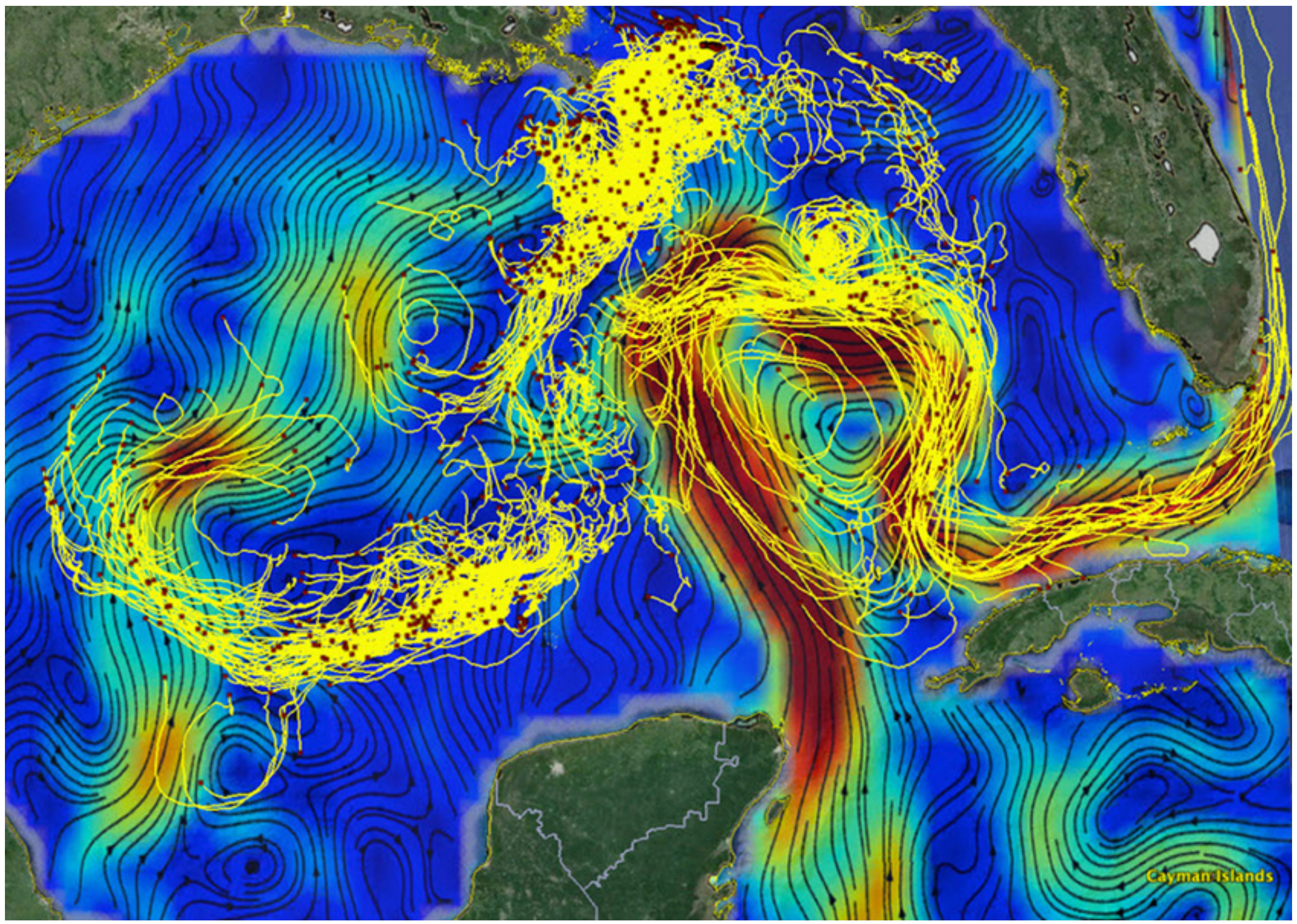
- The ocean current (velocity vector field) varies by space & time
- Scientists get sparse observations of the current from buoys
- Goal: estimate the current “predict”

[Ryan, Özgökmen 2023; Zewe 2023; Gonçalves et al 2019; Lodise et al 2020; Berlinghieri et al 2023]

Challenges:

Why Gaussian processes (GPs)?

- Often want to estimate/“predict” some continuous outcome as a function of certain inputs (*regression*)
- GPs are good at certain types of regression problems



Example:

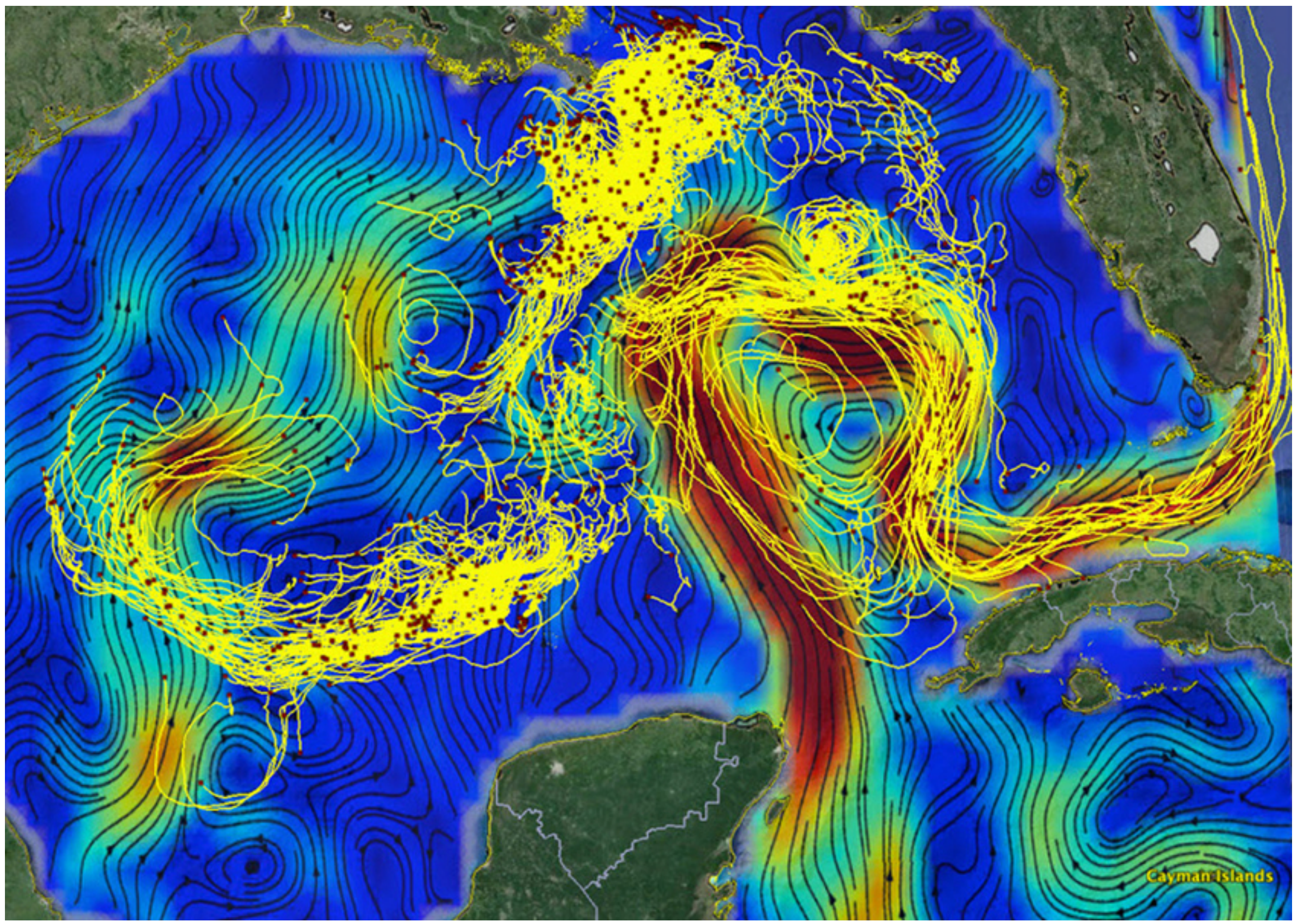
- The ocean current (velocity vector field) varies by space & time
- Scientists get sparse observations of the current from buoys
- Goal: estimate the current “predict”

[Ryan, Özgökmen 2023; Zewe 2023; Gonçalves et al 2019; Lodise et al 2020; Berlinghieri et al 2023]

Challenges: • Sparsely observed data, not on a grid

Why Gaussian processes (GPs)?

- Often want to estimate/“predict” some continuous outcome as a function of certain inputs (*regression*)
- GPs are good at certain types of regression problems



Example:

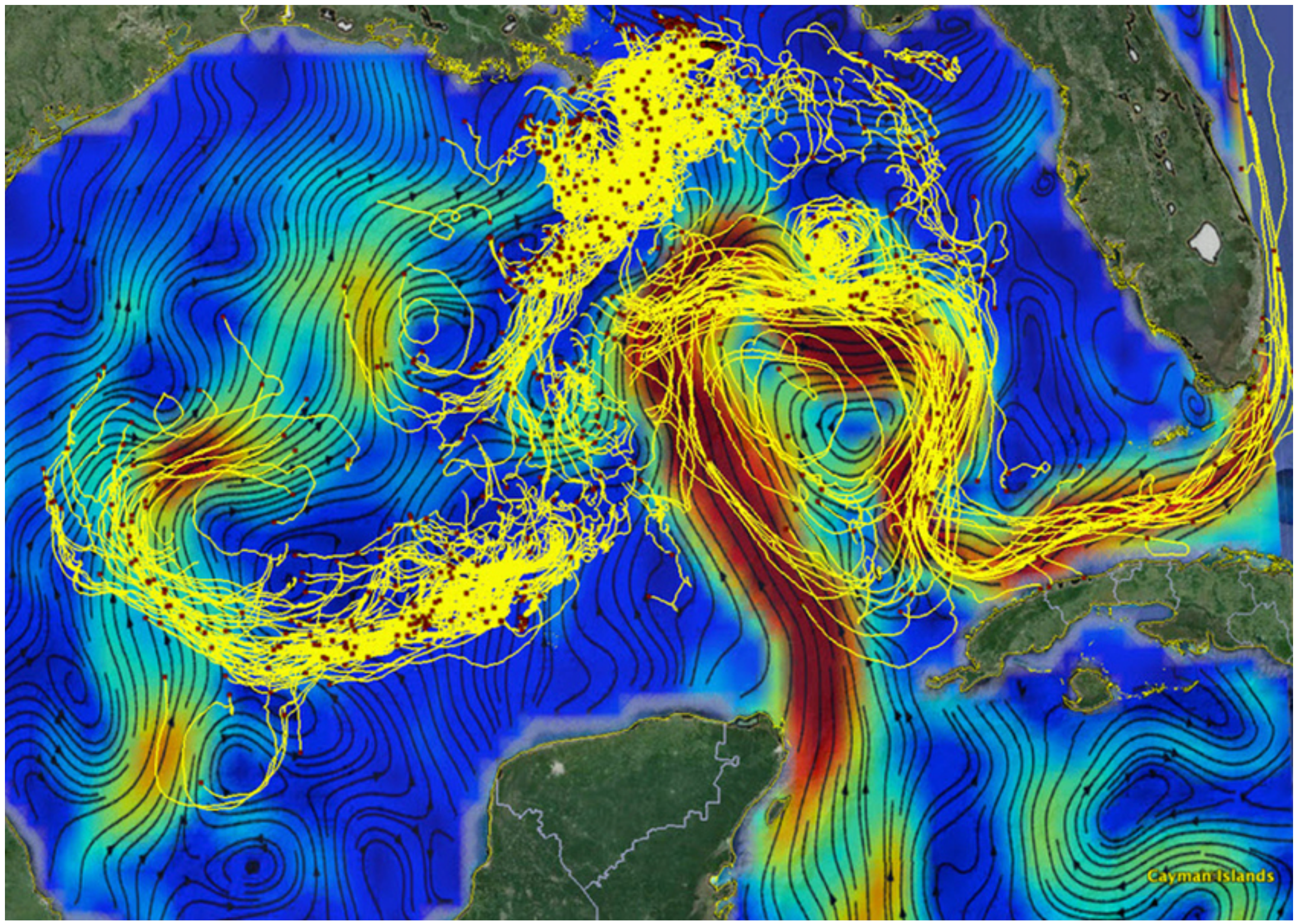
- The ocean current (velocity vector field) varies by space & time
- Scientists get sparse observations of the current from buoys
- Goal: estimate the current “predict”

[Ryan, Özgökmen 2023; Zewe 2023; Gonçalves et al 2019; Lodise et al 2020; Berlinghieri et al 2023]

Challenges:
- Sparsely observed data, not on a grid
- Current is highly nonlinear but smooth in space-time

Why Gaussian processes (GPs)?

- Often want to estimate/"predict" some continuous outcome as a function of certain inputs (*regression*)
- GPs are good at certain types of regression problems



Example:

- The ocean current (velocity vector field) varies by space & time
- Scientists get sparse observations of the current from buoys
- Goal: estimate the current "predict"

[Ryan, Özgökmen 2023; Zewe 2023; Gonçalves et al 2019; Lodise et al 2020; Berlinghieri et al 2023]

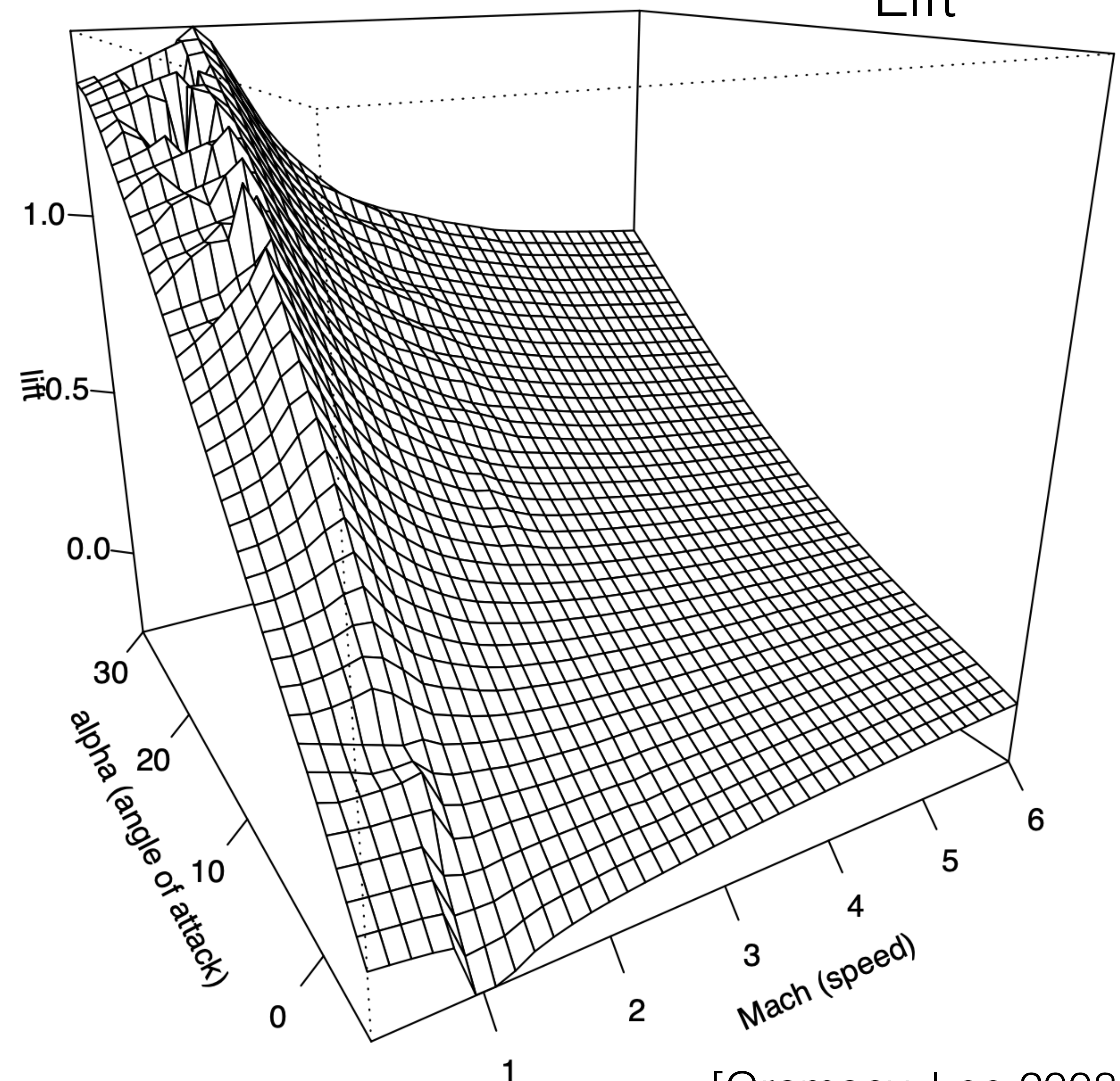
Challenges:
- Sparsely observed data, not on a grid
- Current is highly nonlinear but smooth in space-time
- Want uncertainty quantification

Why Gaussian processes?

Why Gaussian processes?

Example:

Lift



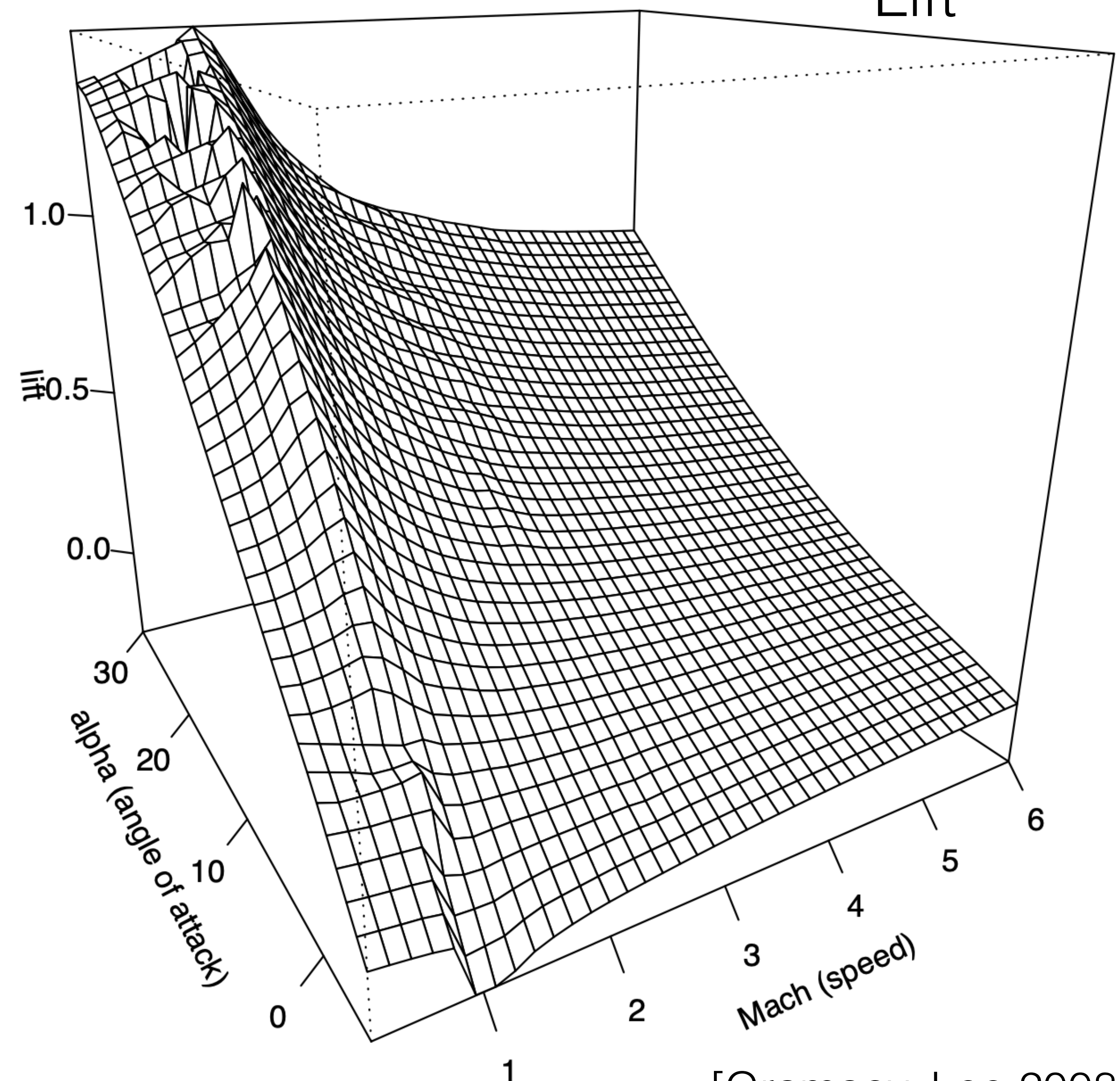
[Gramacy, Lee 2008]
[Gramacy 2020]

Why Gaussian processes?

Example:

- The lift force of a rocket booster varies as a function of speed at re-entry, angle of attack, and sideslip angle

Lift

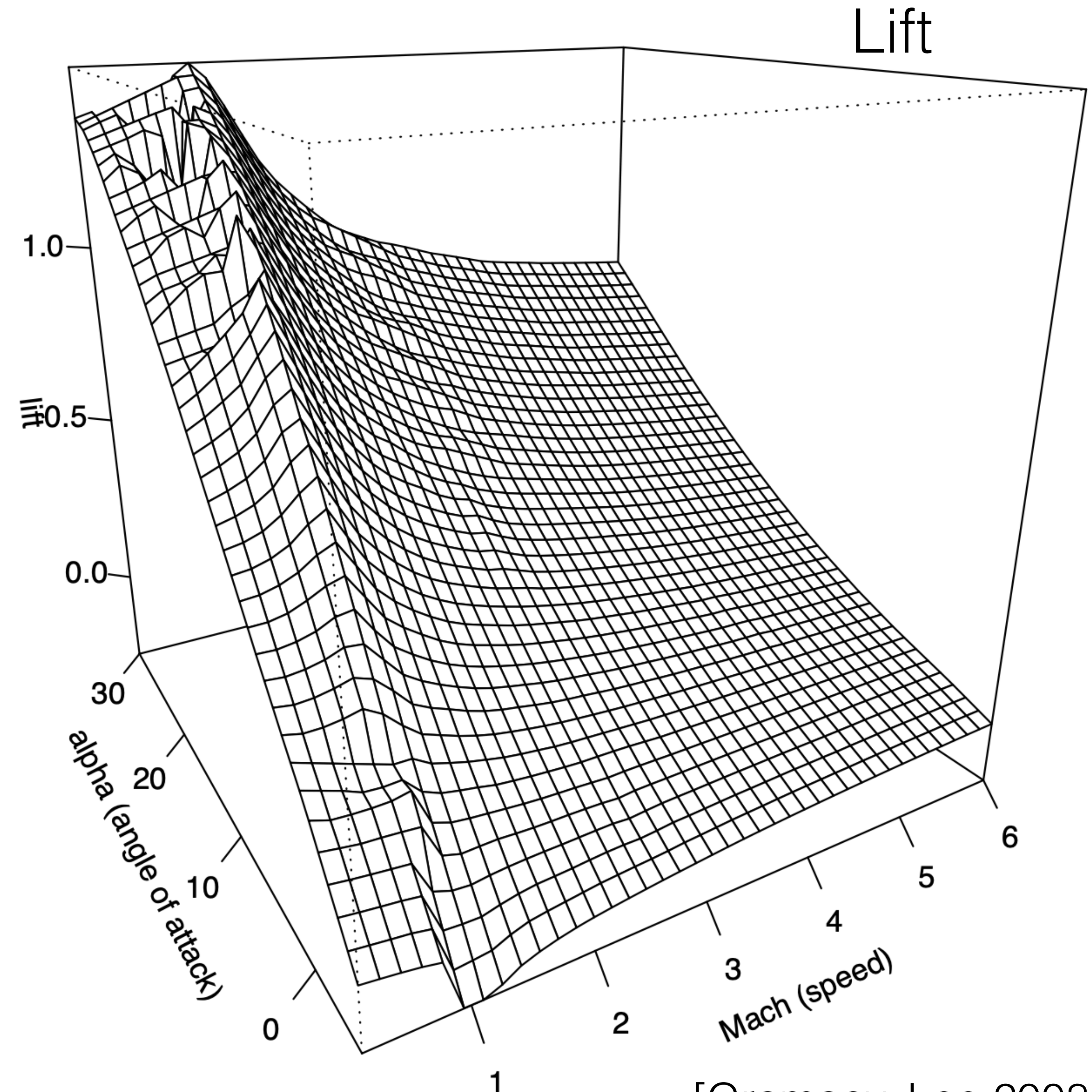


[Gramacy, Lee 2008]
[Gramacy 2020]

Why Gaussian processes?

Example:

- The lift force of a rocket booster varies as a function of speed at re-entry, angle of attack, and sideslip angle
- Scientists can run expensive simulations at chosen input settings

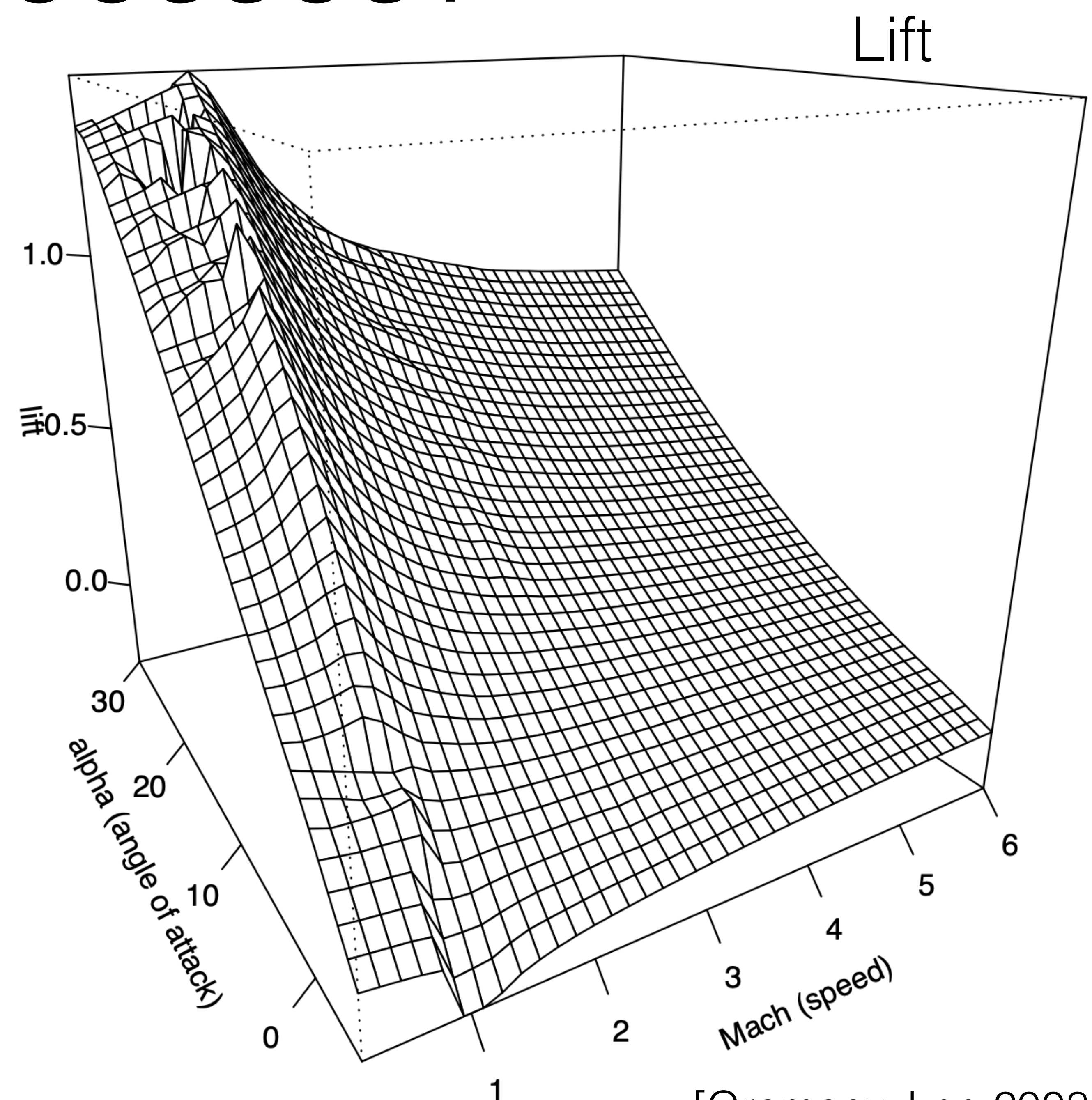


[Gramacy, Lee 2008]
[Gramacy 2020]

Why Gaussian processes?

Example:

- The lift force of a rocket booster varies as a function of speed at re-entry, angle of attack, and sideslip angle
- Scientists can run expensive simulations at chosen input settings
- Goal: estimate how lift varies as a function of these inputs

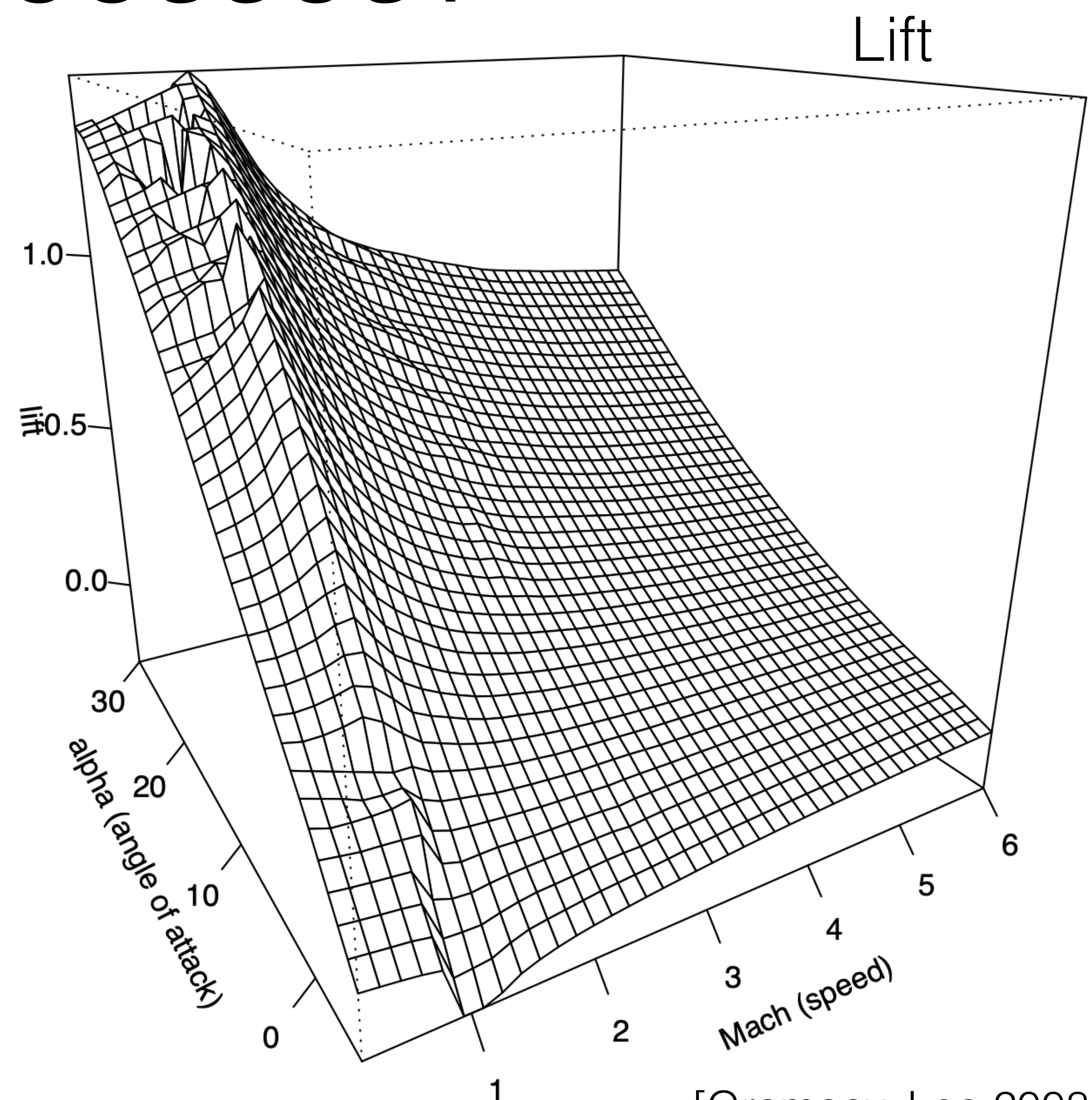


[Gramacy, Lee 2008]
[Gramacy 2020]

Why Gaussian processes?

Example:

- The lift force of a rocket booster varies as a function of speed at re-entry, angle of attack, and sideslip angle
- Scientists can run expensive simulations at chosen input settings
- Goal: estimate how lift varies as a function of these inputs



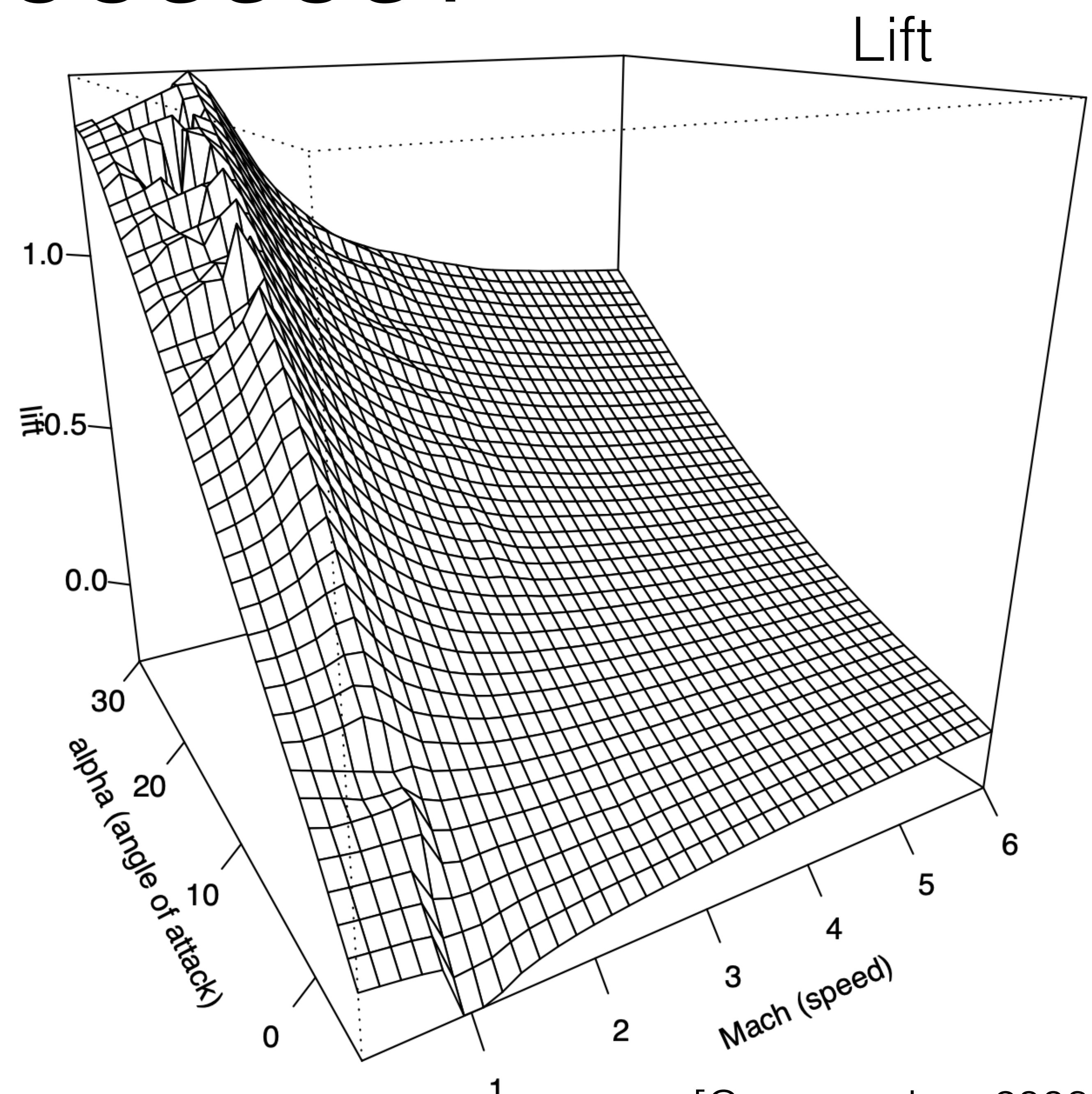
"*Surrogate model*" or "*metamodel*" or "*emulator*"

[Gramacy, Lee 2008]
[Gramacy 2020]

Why Gaussian processes?

Example:

- The lift force of a rocket booster varies as a function of speed at re-entry, angle of attack, and sideslip angle
- Scientists can run expensive simulations at chosen input settings
- Goal: estimate how lift varies as a function of these inputs



"*Surrogate model*" or "*metamodel*" or "*emulator*"

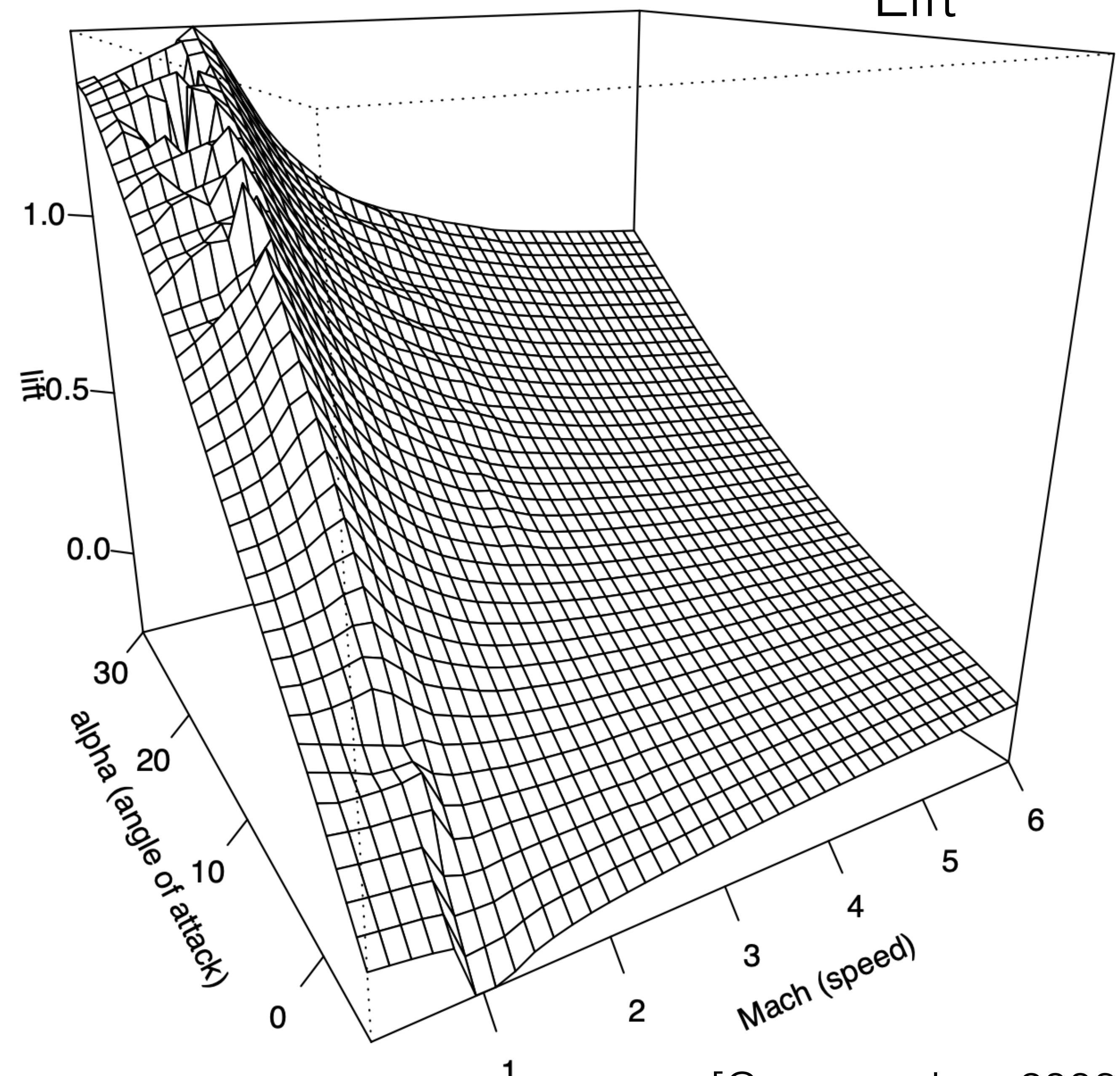
[Gramacy, Lee 2008]
[Gramacy 2020]

Why Gaussian processes?

Example:

- The lift force of a rocket booster varies as a function of speed at re-entry, angle of attack, and sideslip angle
- Scientists can run expensive simulations at chosen input settings
- Goal: estimate how lift varies as a function of these inputs

Lift



"*Surrogate model*" or "*metamodel*" or "*emulator*"

[Gramacy, Lee 2008]
[Gramacy 2020]

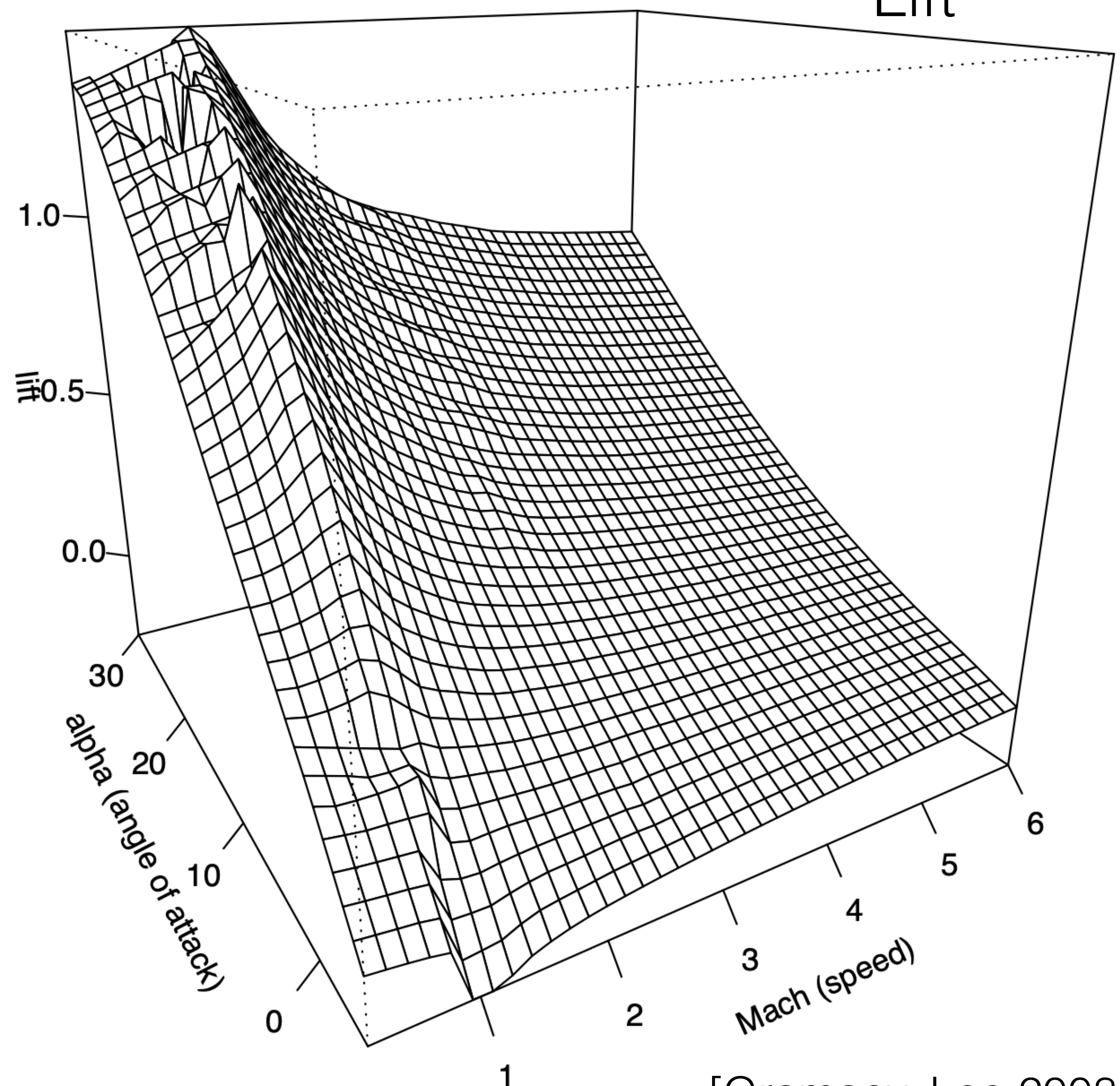
Challenges:

Why Gaussian processes?

Example:

- The lift force of a rocket booster varies as a function of speed at re-entry, angle of attack, and sideslip angle
- Scientists can run expensive simulations at chosen input settings
- Goal: estimate how lift varies as a function of these inputs

Lift



"*Surrogate model*" or "*metamodel*" or "*emulator*"

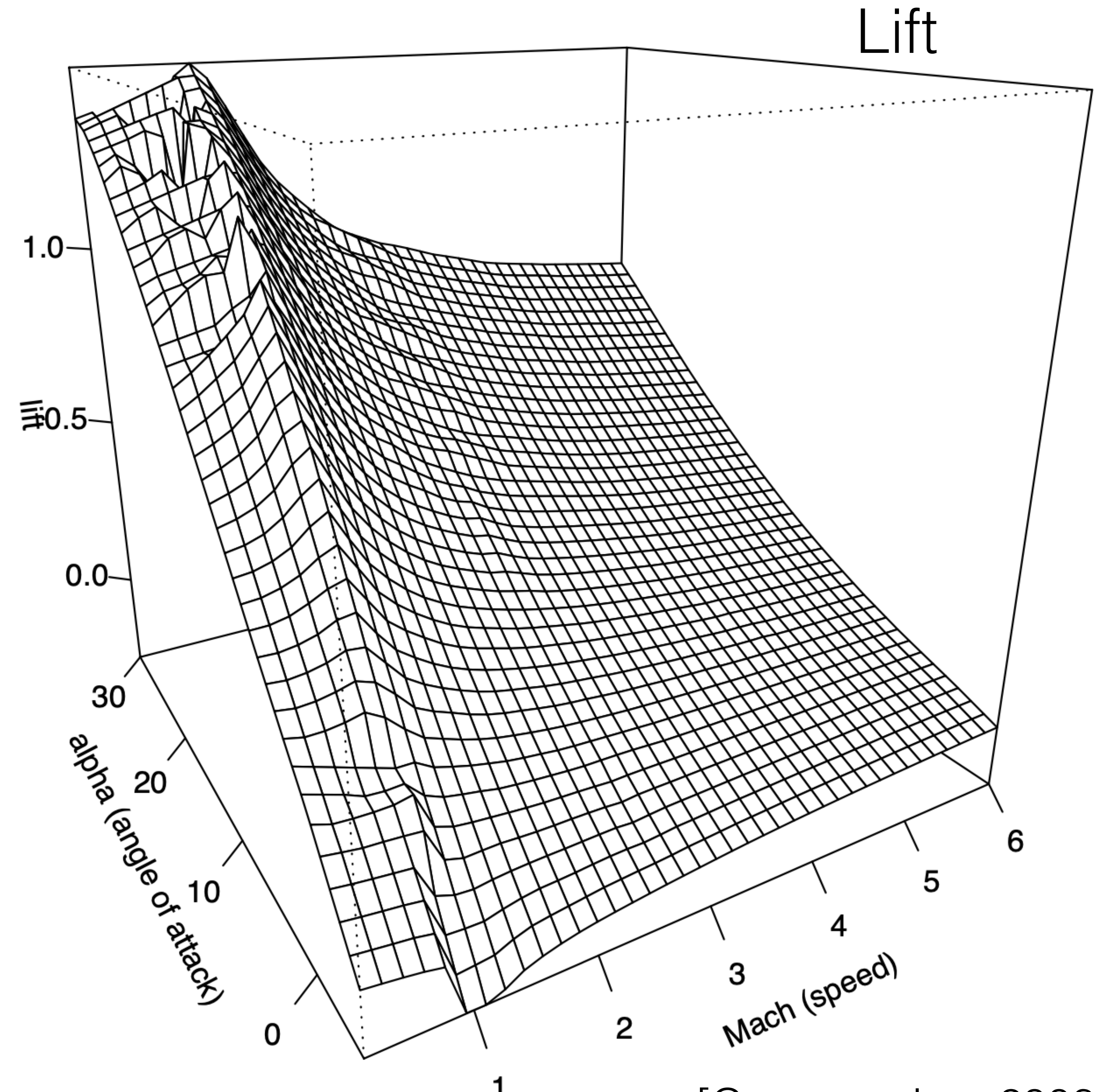
[Gramacy, Lee 2008]
[Gramacy 2020]

Challenges: • Sparsely observed data

Why Gaussian processes?

Example:

- The lift force of a rocket booster varies as a function of speed at re-entry, angle of attack, and sideslip angle
- Scientists can run expensive simulations at chosen input settings
- Goal: estimate how lift varies as a function of these inputs



"*Surrogate model*" or "*metamodel*" or "*emulator*"

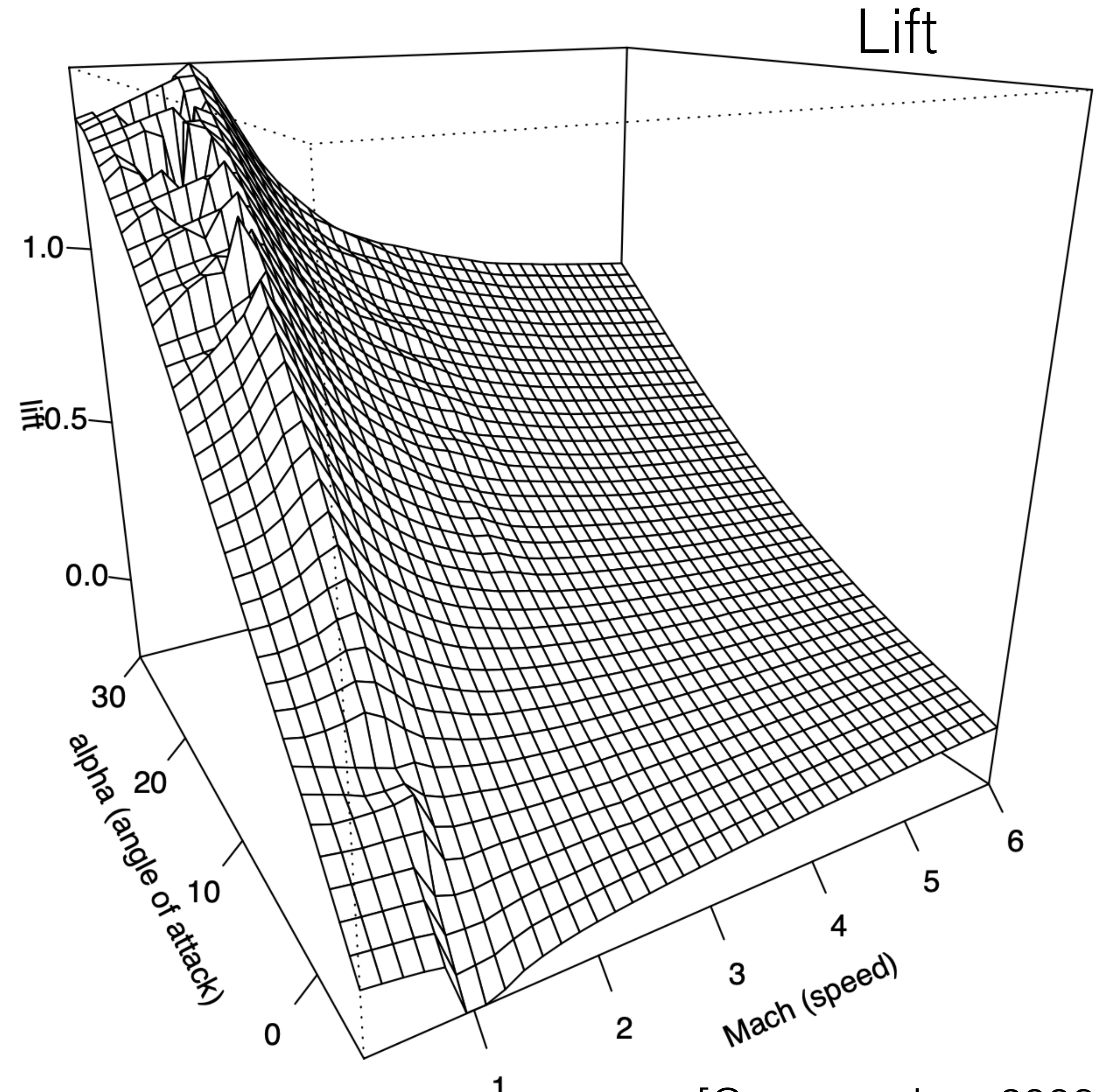
[Gramacy, Lee 2008]
[Gramacy 2020]

Challenges:
- Sparsely observed data
- Lift may have a nonlinear relationship to the inputs

Why Gaussian processes?

Example:

- The lift force of a rocket booster varies as a function of speed at re-entry, angle of attack, and sideslip angle
- Scientists can run expensive simulations at chosen input settings
- Goal: estimate how lift varies as a function of these inputs



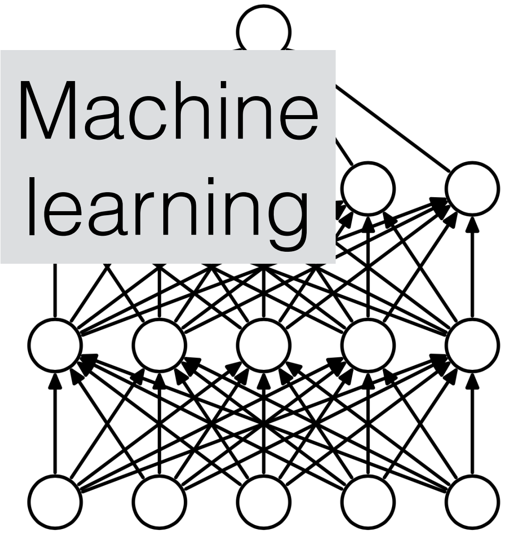
"*Surrogate model*" or "*metamodel*" or "*emulator*"

[Gramacy, Lee 2008]
[Gramacy 2020]

Challenges:
- Sparsely observed data
- Lift may have a nonlinear relationship to the inputs
- Want uncertainty quantification

Why Gaussian processes?

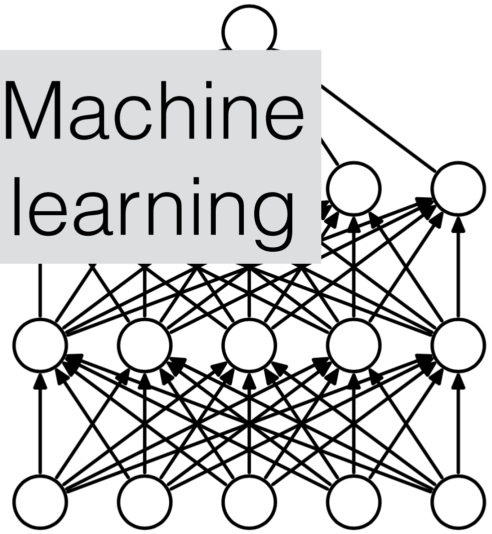
Why Gaussian processes?



[Murphy 2022, 13.18; Snoek et al 2012; Li et al 2023]

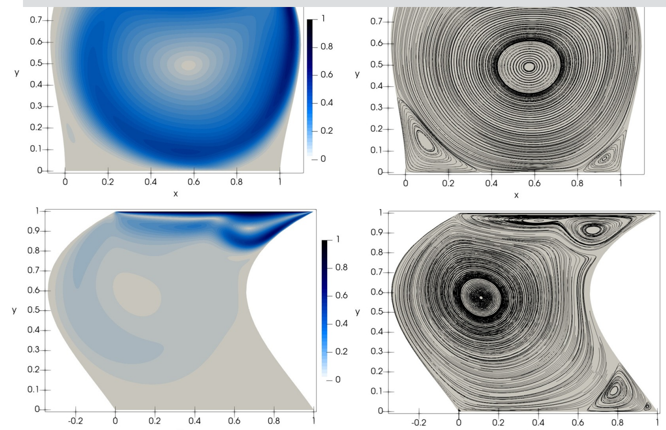
Why Gaussian processes?

Machine learning



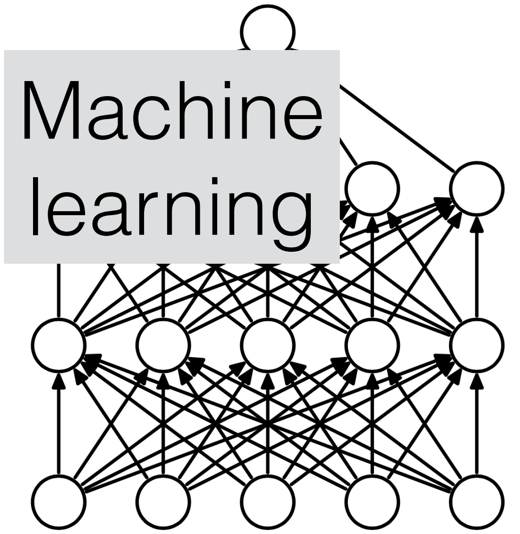
[Murphy 2022, 13.18; Snoek et al 2012; Li et al 2023]

Computational Fluid Dynamics

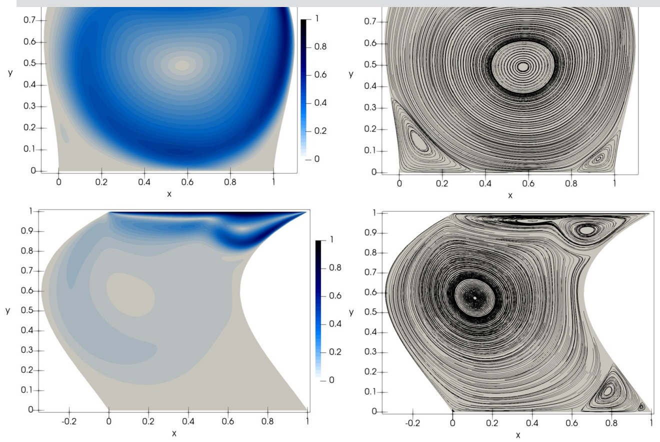


[Morita et al 2022]

Why Gaussian processes?

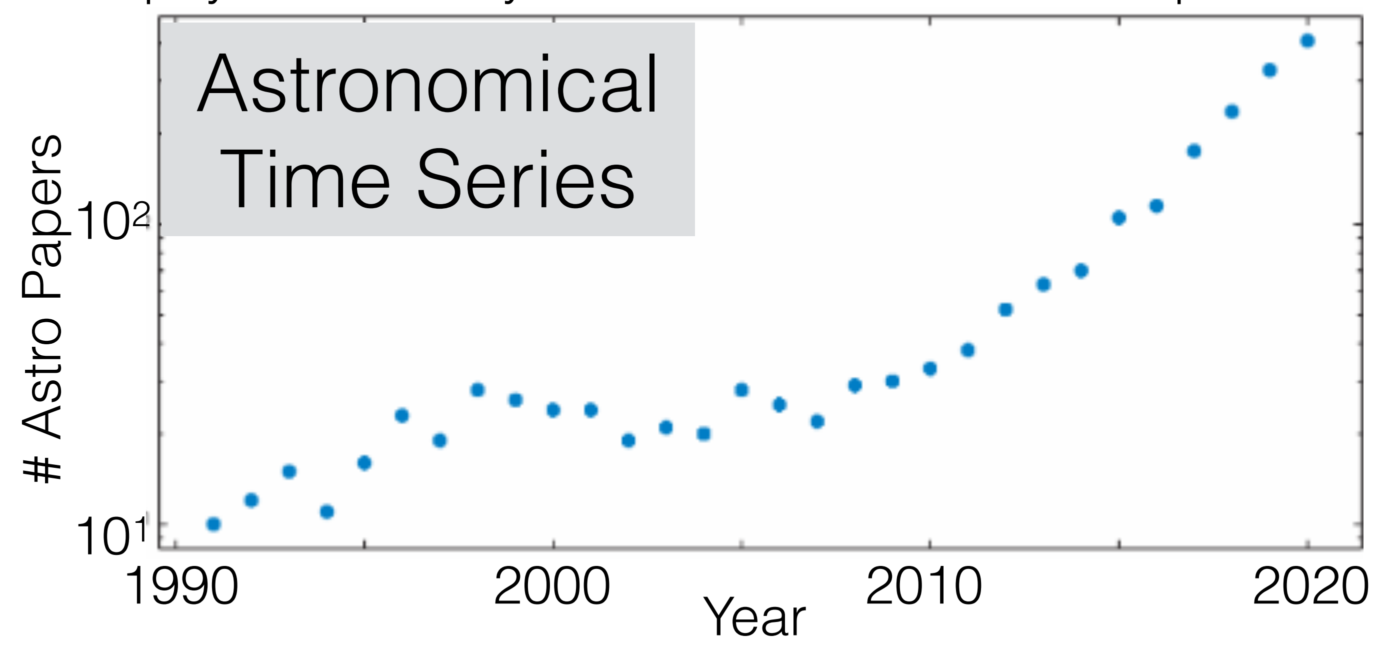


[Murphy 2022, 13.18;
Snoek et al 2012;
Li et al 2023]



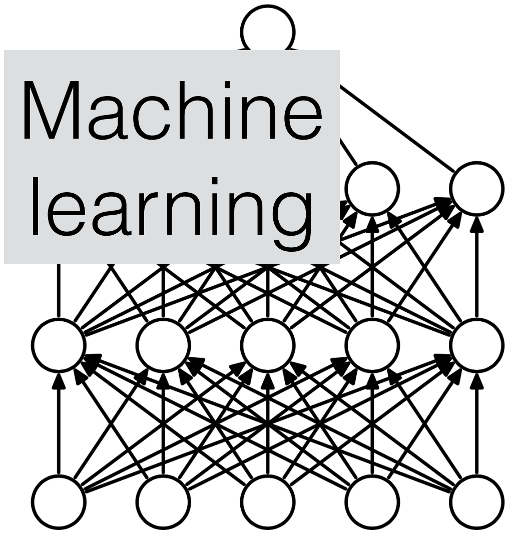
[Morita et al 2022]

Astrophysics Data System search for “Gaussian process”

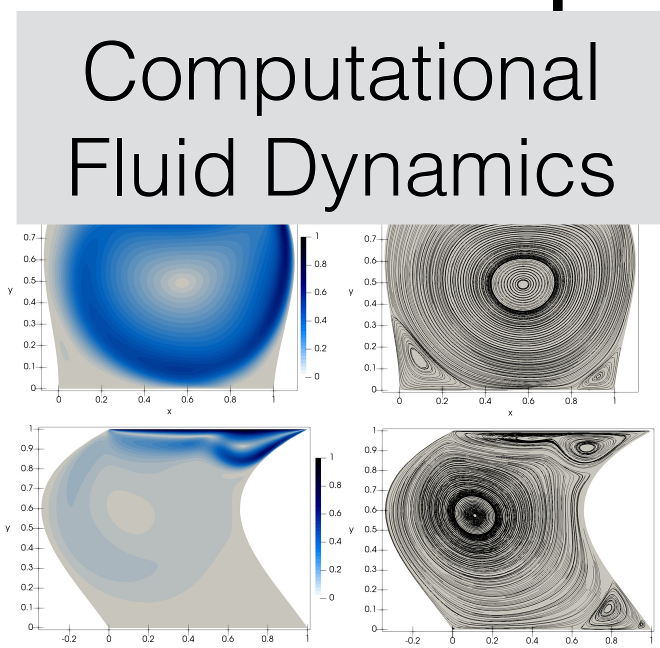


[Aigrain, Foreman-Mackey 2023]

Why Gaussian processes?

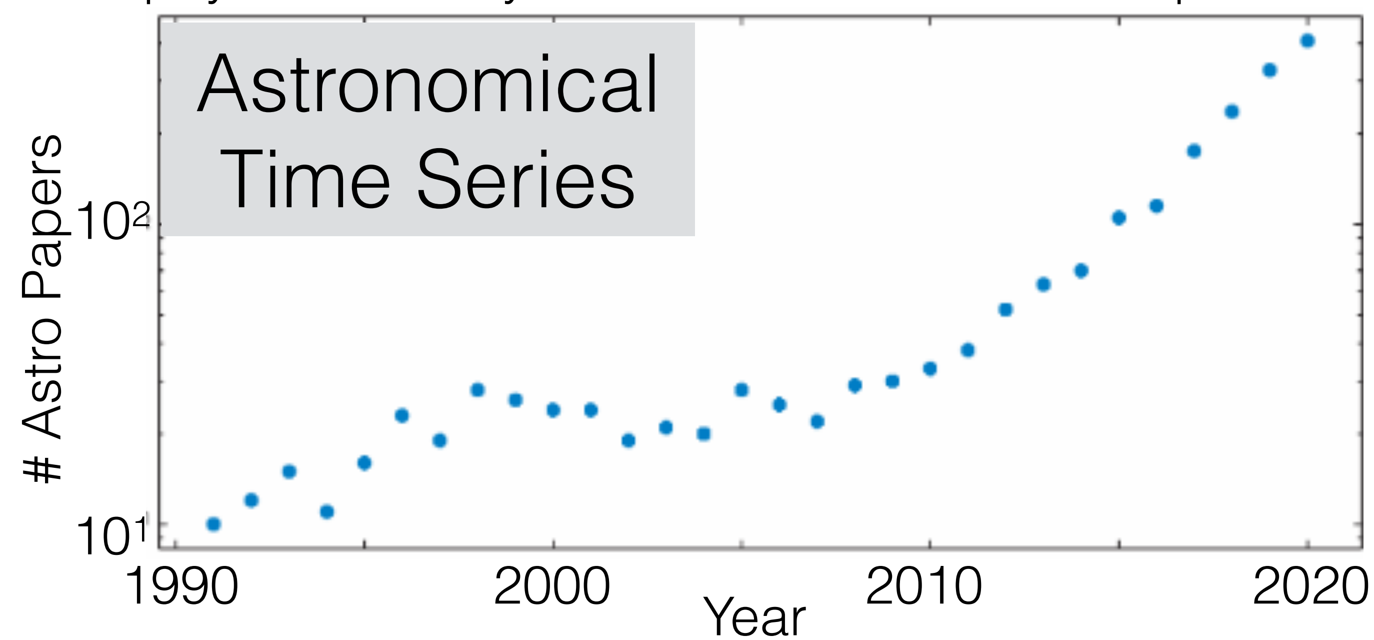


[Murphy 2022, 13.18;
Snoek et al 2012;
Li et al 2023]



[Morita et al 2022]

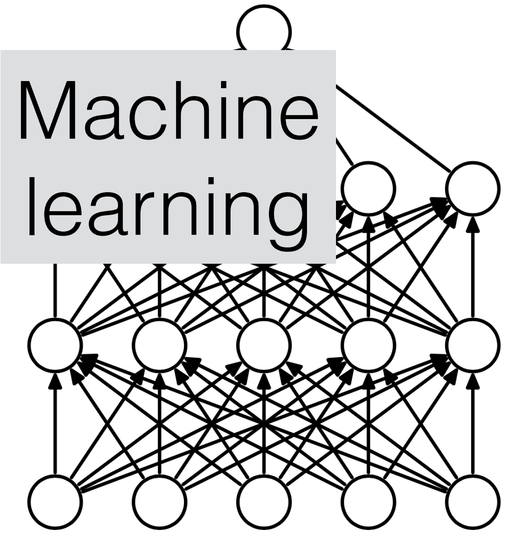
Astrophysics Data System search for "Gaussian process"



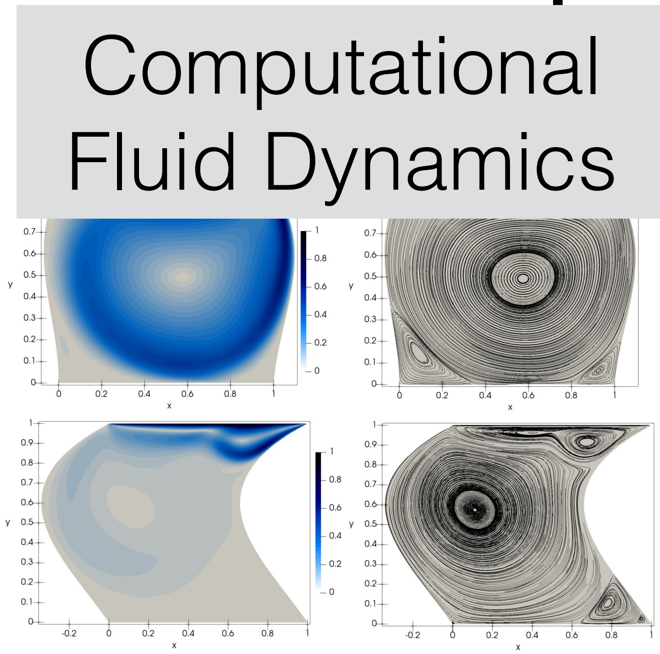
[Aigrain, Foreman-Mackey 2023]

A recurring motif:

Why Gaussian processes?

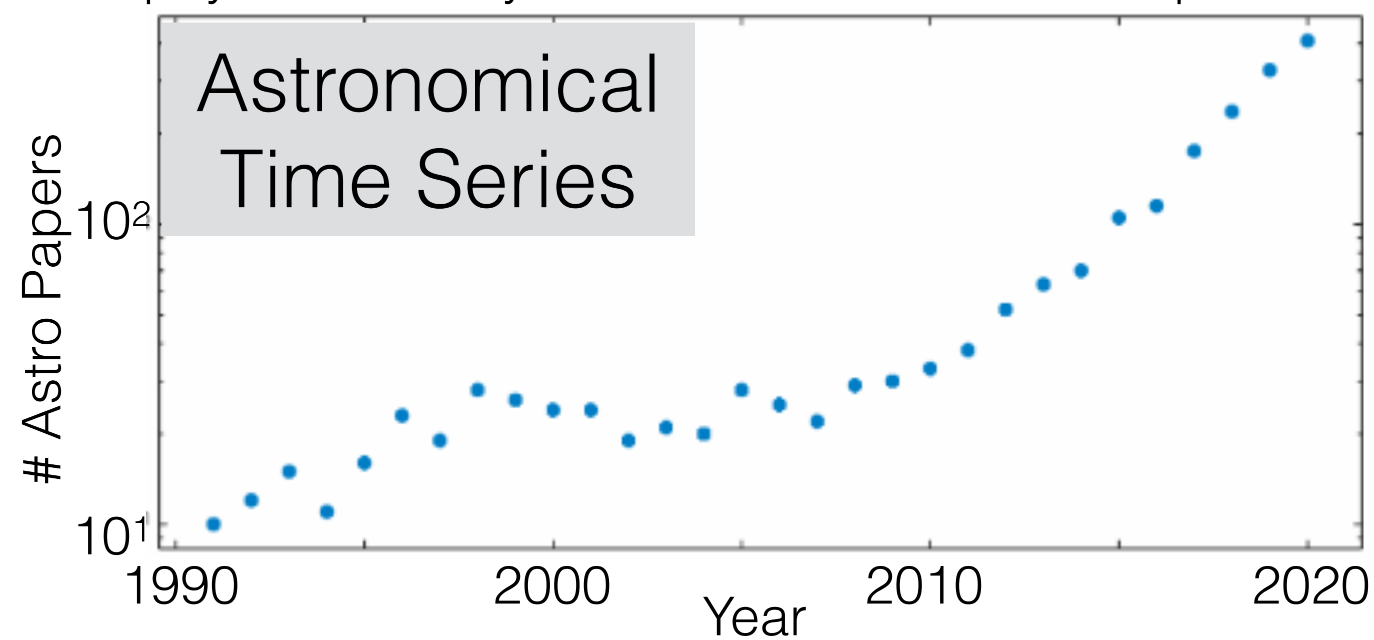


[Murphy 2022, 13.18; Snoek et al 2012; Li et al 2023]



[Morita et al 2022]

Astrophysics Data System search for "Gaussian process"

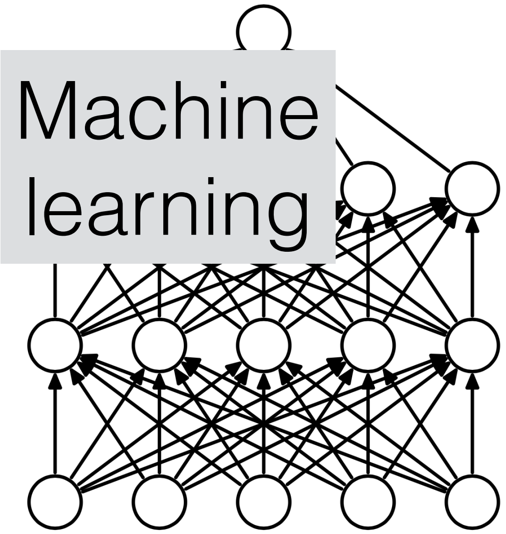


[Aigrain, Foreman-Mackey 2023]

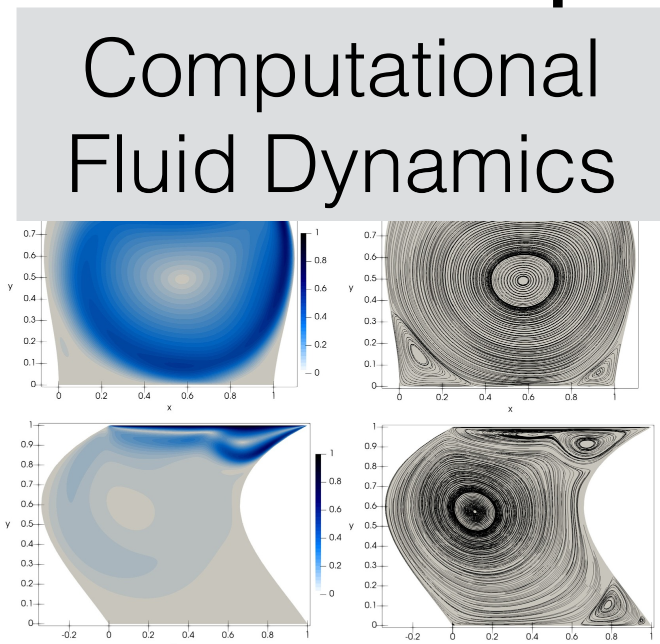
A recurring motif:

- We'd like to estimate a potentially nonlinear but smooth function of a handful of inputs

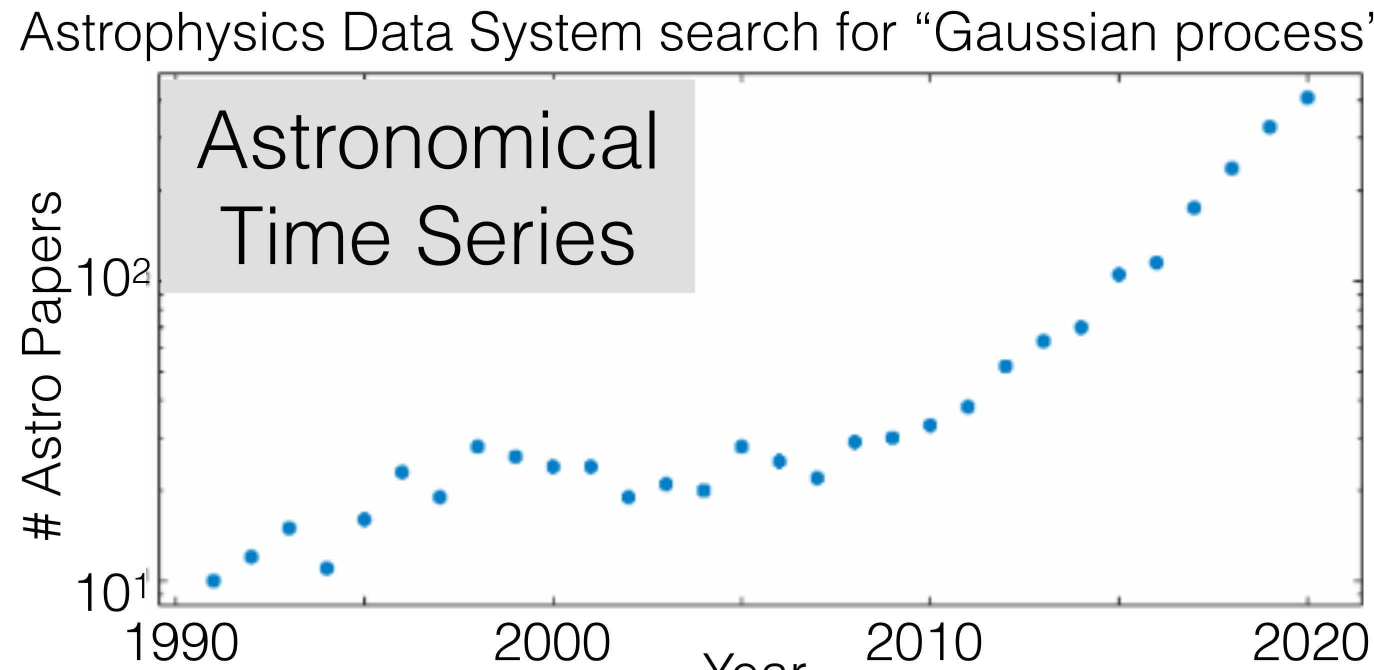
Why Gaussian processes?



[Murphy 2022, 13.18; Snoek et al 2012; Li et al 2023]



[Morita et al 2022]

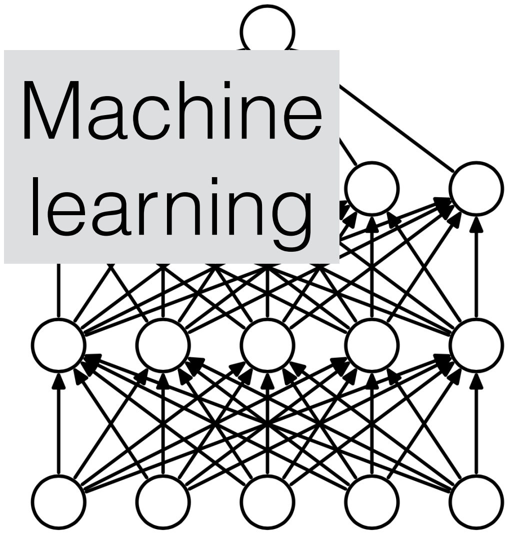


[Aigrain, Foreman-Mackey 2023]

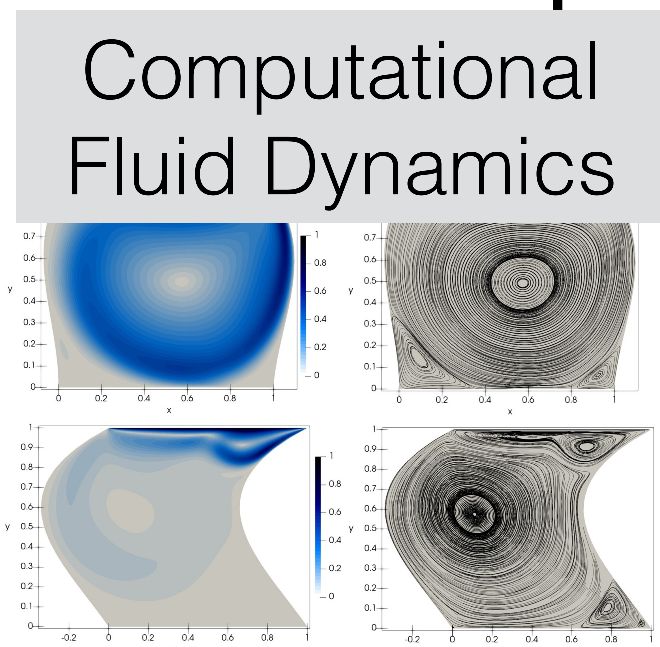
A recurring motif:

- We'd like to estimate a potentially nonlinear but smooth function of a handful of inputs
- We have access to sparse (possibly noisy) observations of the function

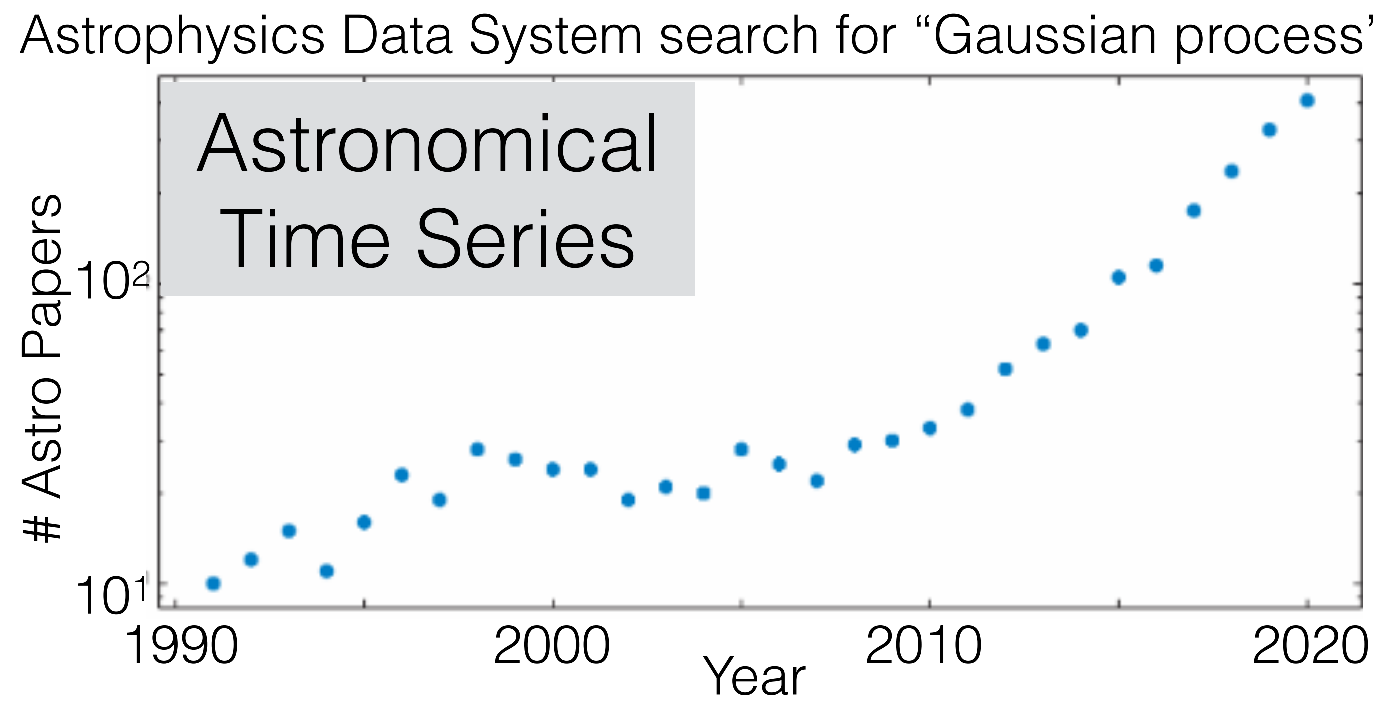
Why Gaussian processes?



[Murphy 2022, 13.18;
Snoek et al 2012;
Li et al 2023]



[Morita et al 2022]

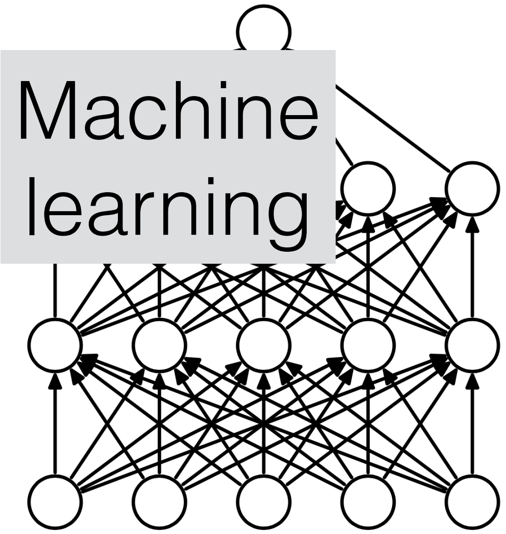


[Aigrain, Foreman-Mackey 2023]

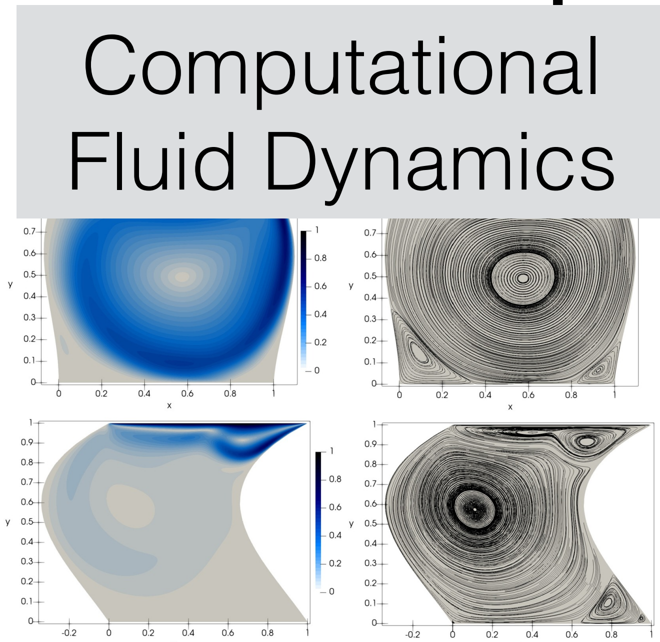
A recurring motif:

- We'd like to estimate a potentially nonlinear but smooth function of a handful of inputs
- We have access to sparse (possibly noisy) observations of the function
- We'd like to quantify our uncertainty

Why Gaussian processes?

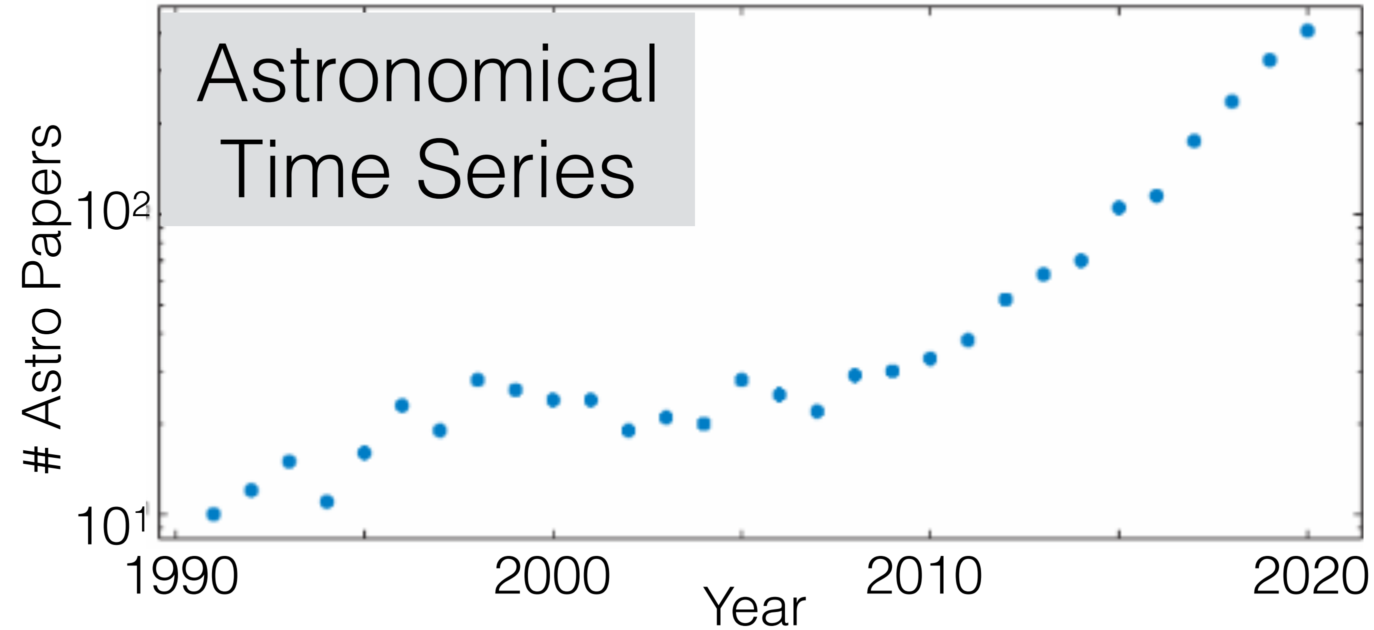


[Murphy 2022, 13.18; Snoek et al 2012; Li et al 2023]



[Morita et al 2022]

Astrophysics Data System search for “Gaussian process”



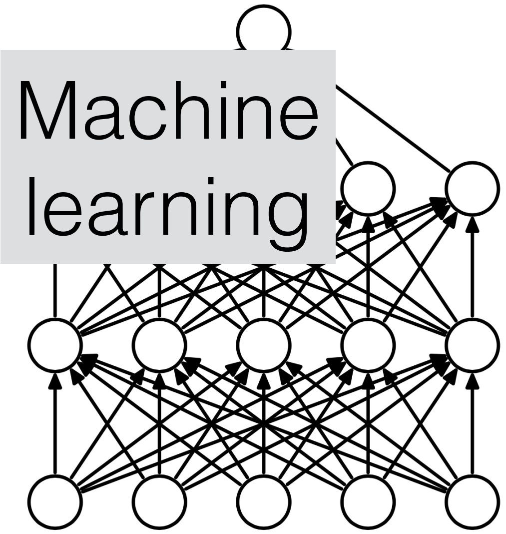
[Aigrain, Foreman-Mackey 2023]

A recurring motif:

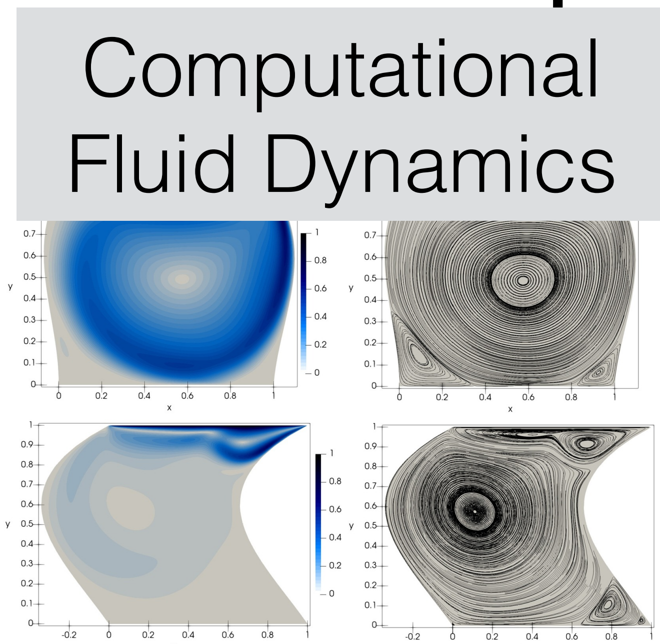
- We’d like to estimate a potentially nonlinear but smooth function of a handful of inputs
- We have access to sparse (possibly noisy) observations of the function
- We’d like to quantify our uncertainty

Bonus benefits:

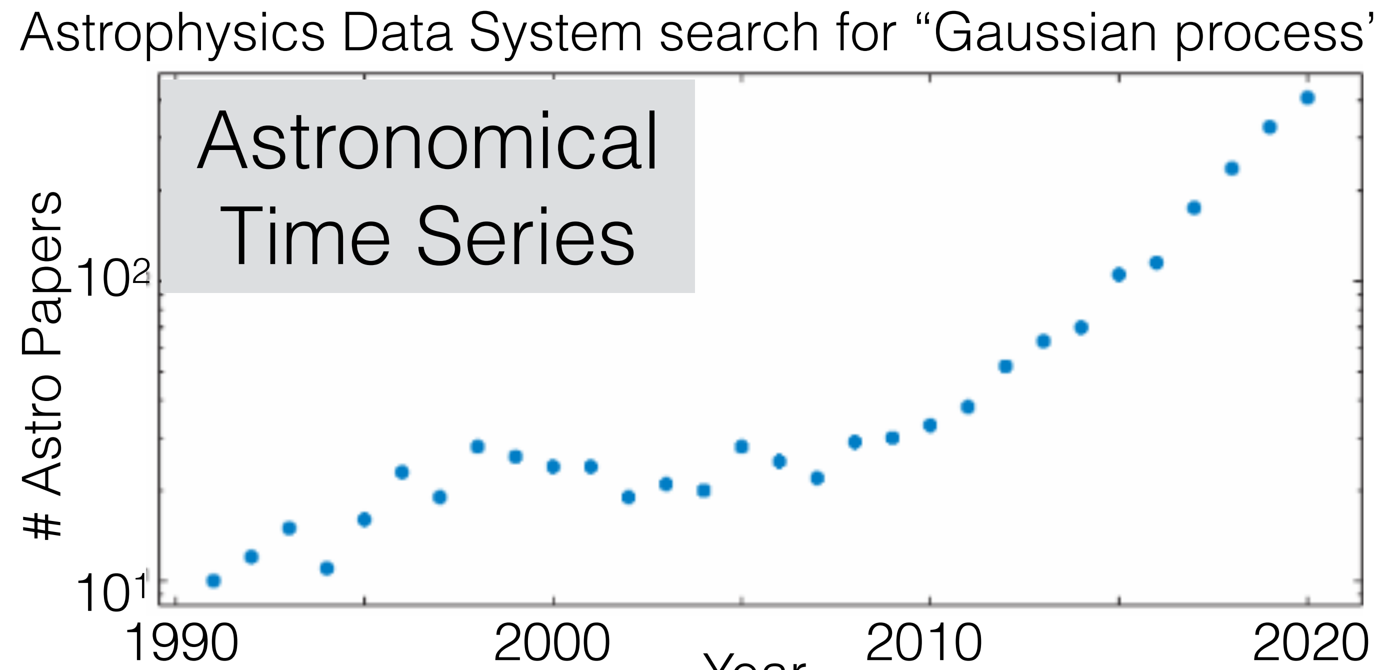
Why Gaussian processes?



[Murphy 2022, 13.18; Snoek et al 2012; Li et al 2023]



[Morita et al 2022]



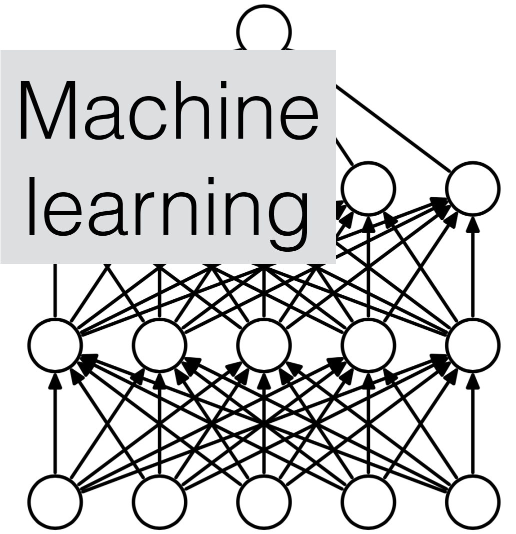
[Aigrain, Foreman-Mackey 2023]

A recurring motif:

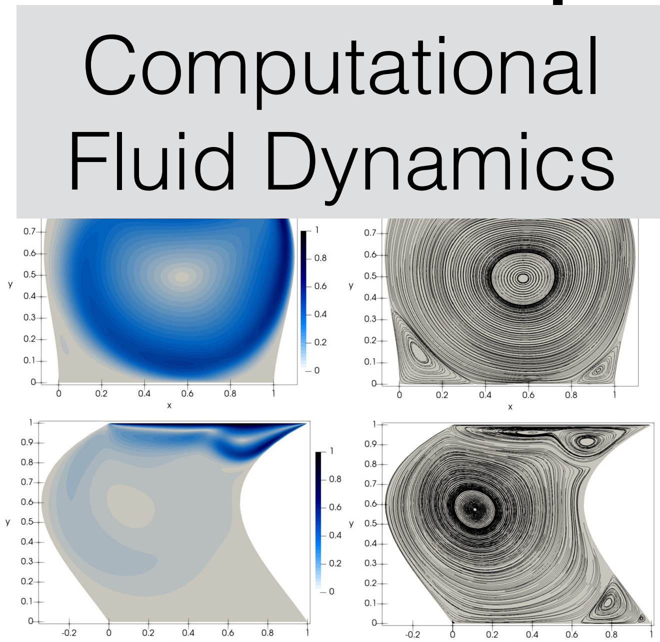
- We’d like to estimate a potentially nonlinear but smooth function of a handful of inputs
- We have access to sparse (possibly noisy) observations of the function
- We’d like to quantify our uncertainty

Bonus benefits: • Ease of use (software, tuning)

Why Gaussian processes?

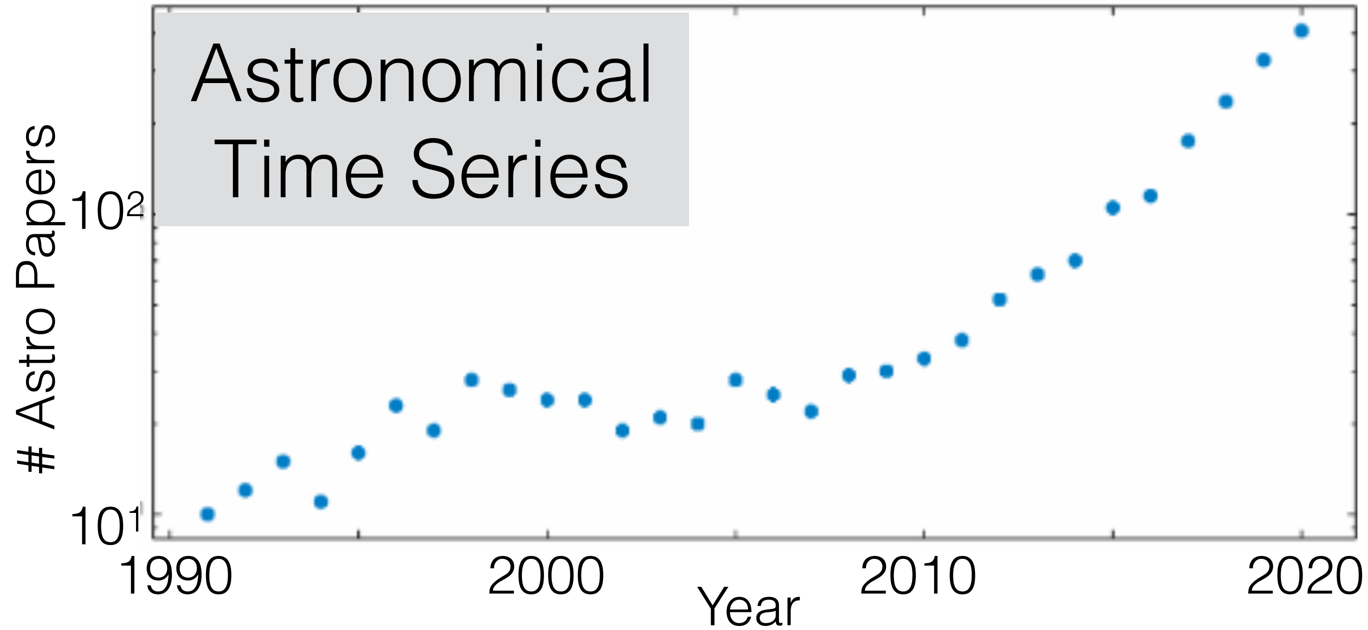


[Murphy 2022, 13.18;
Snoek et al 2012;
Li et al 2023]



[Morita et al 2022]

Astrophysics Data System search for "Gaussian process"



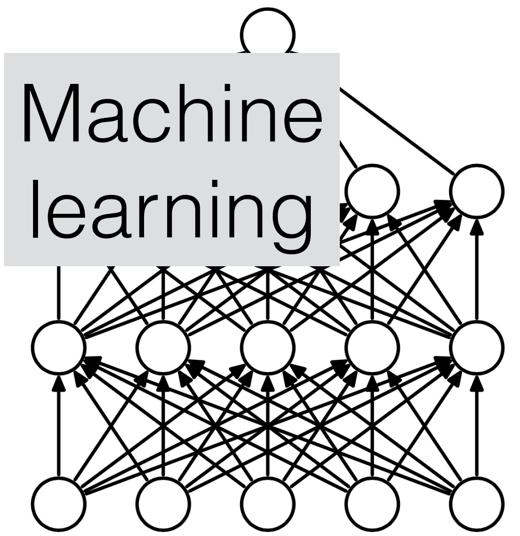
[Aigrain, Foreman-Mackey 2023]

A recurring motif:

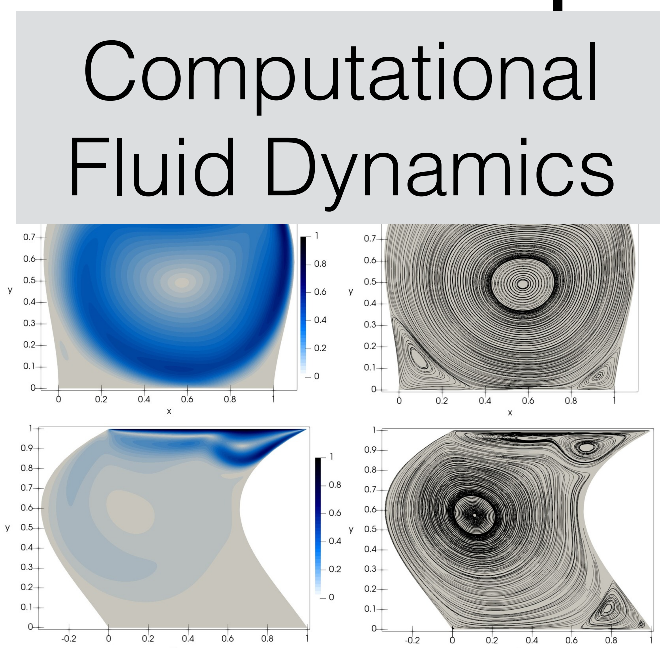
- We'd like to estimate a potentially nonlinear but smooth function of a handful of inputs
- We have access to sparse (possibly noisy) observations of the function
- We'd like to quantify our uncertainty

Bonus benefits: • Ease of use (software, tuning) • Supports optimization of outcome

Why Gaussian processes?

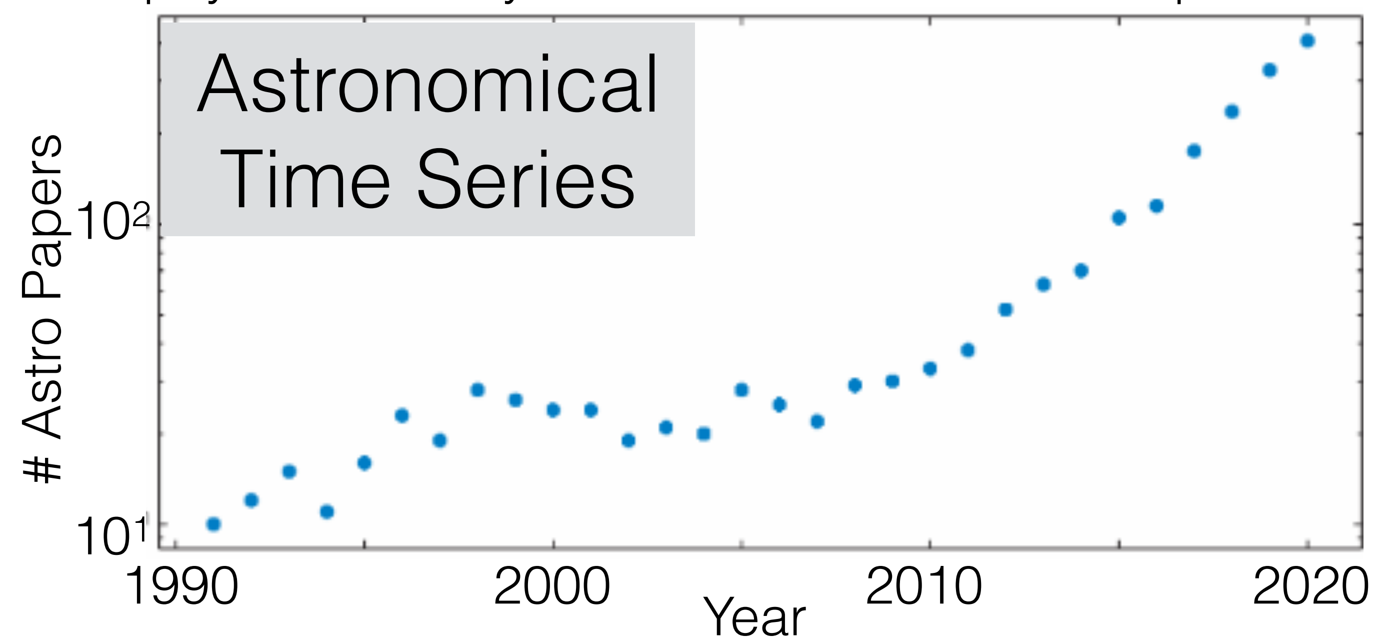


[Murphy 2022, 13.18;
Snoek et al 2012;
Li et al 2023]



[Morita et al 2022]

Astrophysics Data System search for “Gaussian process”



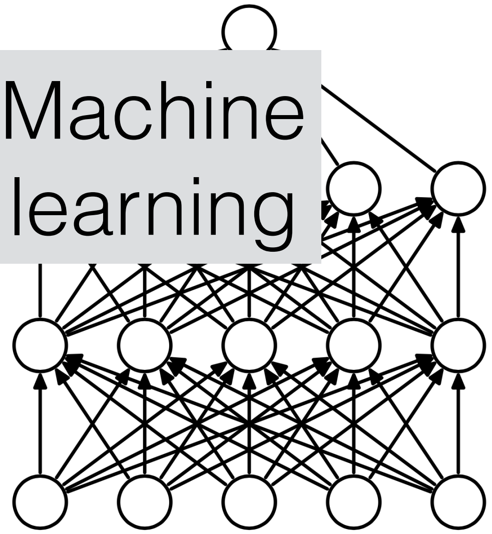
[Aigrain, Foreman-Mackey 2023]

A recurring motif:

- We’d like to estimate a potentially nonlinear but smooth function of a handful of inputs
- We have access to sparse (possibly noisy) observations of the function
- We’d like to quantify our uncertainty

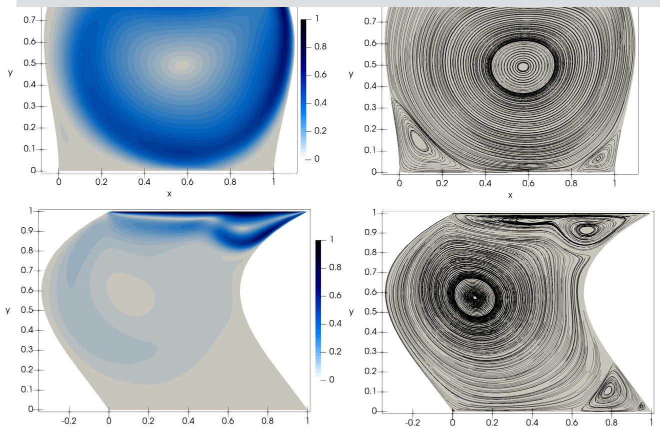
Bonus benefits: • Ease of use (software, tuning) • Supports optimization of outcome • Predictions & uncertainties over derivatives & integrals

Why Gaussian processes?



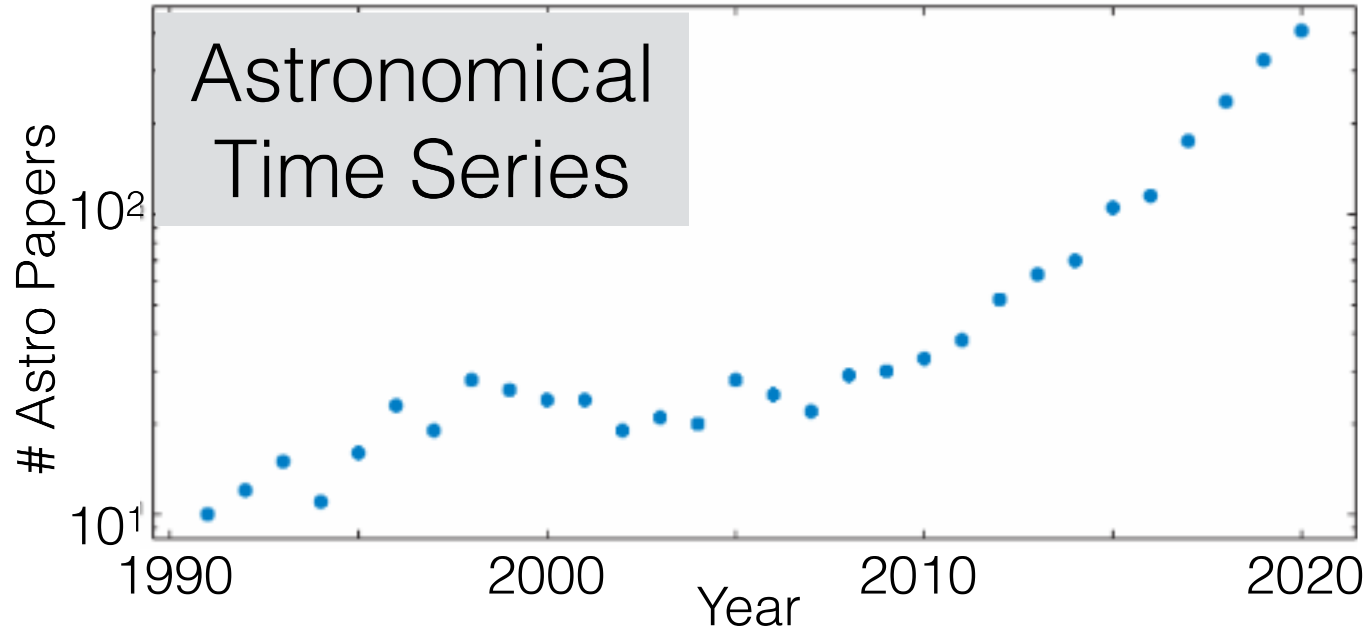
[Murphy 2022, 13.18;
Snoek et al 2012;
Li et al 2023]

Computational
Fluid Dynamics



[Morita et al 2022]

Astrophysics Data System search for “Gaussian process”



[Aigrain, Foreman-Mackey 2023]

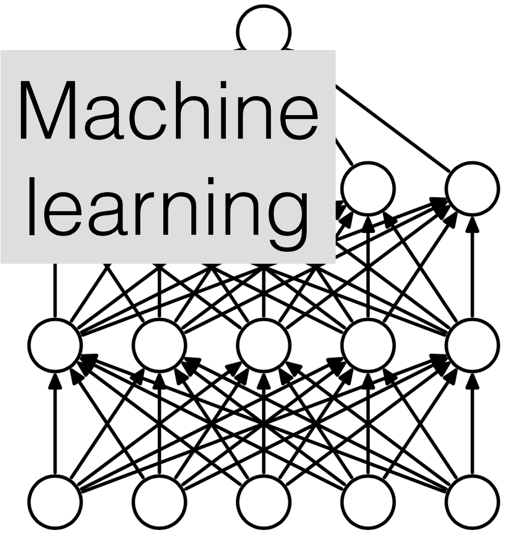
A recurring motif:

- We’d like to estimate a potentially nonlinear but smooth function of a handful of inputs
- We have access to sparse (possibly noisy) observations of the function
- We’d like to quantify our uncertainty

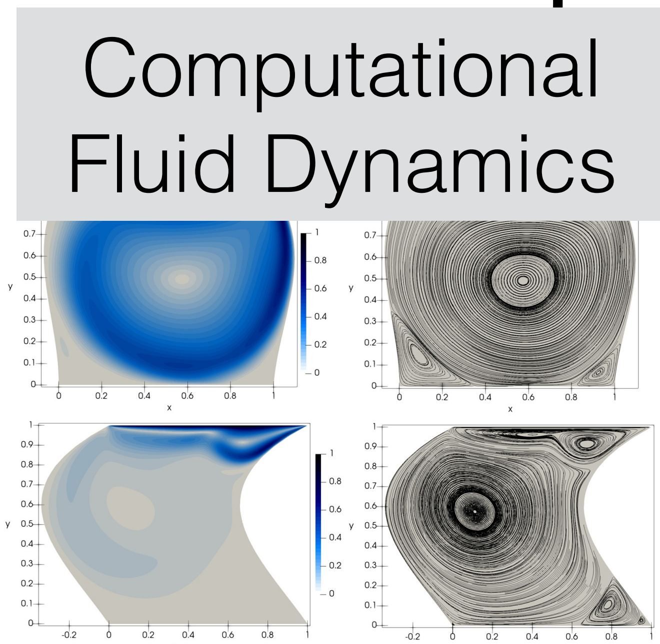
Bonus benefits: • Ease of use (software, tuning) • Supports optimization of outcome • Predictions & uncertainties over derivatives & integrals • Module in more-complex methods

Why Gaussian processes?

see also "kriging," "optimal interpolation (OI)"

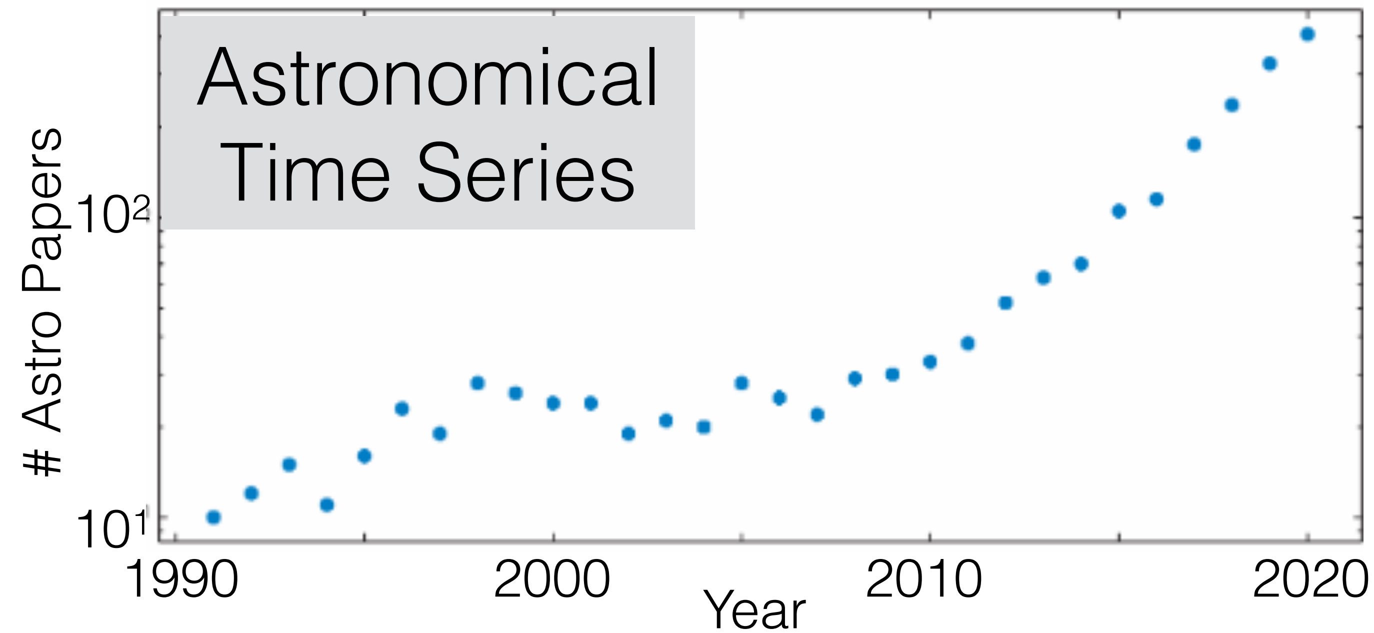


[Murphy 2022, 13.18; Snoek et al 2012; Li et al 2023]



[Morita et al 2022]

Astrophysics Data System search for "Gaussian process"



[Aigrain, Foreman-Mackey 2023]

A recurring motif:

- We'd like to estimate a potentially nonlinear but smooth function of a handful of inputs
- We have access to sparse (possibly noisy) observations of the function
- We'd like to quantify our uncertainty

Bonus benefits: • Ease of use (software, tuning) • Supports optimization of outcome • Predictions & uncertainties over derivatives & integrals • Module in more-complex methods

Roadmap

- Bayesian modeling and inference
- Gaussian process model
  - Popular version using a squared exponential kernel
- Gaussian process inference
  - Prediction & uncertainty quantification
- Observation noise
- What uncertainty are we quantifying?
- What can go wrong?
- Bayesian optimization

- Goals:
  - Learn the mechanism behind standard GPs to identify benefits and pitfalls (also in BayesOpt)
  - Learn the skills to be responsible users of standard GPs (transferable to other ML/AI methods)

Roadmap

- Bayesian modeling and inference
- Gaussian process model
  - Popular version using a squared exponential kernel
- Gaussian process inference
  - Prediction & uncertainty quantification
- Observation noise
- What uncertainty are we quantifying?
- What can go wrong?
- Bayesian optimization
- Goals:
  - Learn the mechanism behind standard GPs to identify benefits and pitfalls (also in BayesOpt)
  - Learn the skills to be responsible users of standard GPs (transferable to other ML/AI methods)

Roadmap

- Bayesian modeling and inference
- Gaussian process model
  - Popular version using a squared exponential kernel
- Gaussian process inference
  - Prediction & uncertainty quantification
- Observation noise
- What uncertainty are we quantifying?
- What can go wrong?
- Bayesian optimization

- Goals:
  - Learn the mechanism behind standard GPs to identify benefits and pitfalls (also in BayesOpt)
  - Learn the skills to be responsible users of standard GPs (transferable to other ML/AI methods)

Roadmap

- Bayesian modeling and inference
- Gaussian process model
  - Popular version using a squared exponential kernel
- Gaussian process inference
  - Prediction & uncertainty quantification
- Observation noise
- What uncertainty are we quantifying?
- What can go wrong?
- Bayesian optimization
- Goals:
  - Learn the mechanism behind standard GPs to identify benefits and pitfalls (also in BayesOpt)
  - Learn the skills to be responsible users of standard GPs (transferable to other ML/AI methods)

Roadmap

- Bayesian modeling and inference
- Gaussian process model
  - Popular version using a squared exponential kernel
- Gaussian process inference
  - Prediction & uncertainty quantification
- Observation noise
- What uncertainty are we quantifying?
- What can go wrong?
- Bayesian optimization

- Goals:
  - Learn the mechanism behind standard GPs to identify benefits and pitfalls (also in BayesOpt)
  - Learn the skills to be responsible users of standard GPs (transferable to other ML/AI methods)

Roadmap

- Bayesian modeling and inference
- Gaussian process model
  - Popular version using a squared exponential kernel
- Gaussian process inference
  - Prediction & uncertainty quantification
- Observation noise
- What uncertainty are we quantifying?
- What can go wrong?
- Bayesian optimization

- Goals:
  - Learn the mechanism behind standard GPs to identify benefits and pitfalls (also in BayesOpt)
  - Learn the skills to be responsible users of standard GPs (transferable to other ML/AI methods)

Roadmap

- Bayesian modeling and inference
- Gaussian process model
  - Popular version using a squared exponential kernel
- Gaussian process inference
  - Prediction & uncertainty quantification
- Observation noise
- What uncertainty are we quantifying?
- What can go wrong?
- Bayesian optimization

- Goals:
  - Learn the mechanism behind standard GPs to identify benefits and pitfalls (also in BayesOpt)
  - Learn the skills to be responsible users of standard GPs (transferable to other ML/AI methods)

Roadmap

- Bayesian modeling and inference
- Gaussian process model
  - Popular version using a squared exponential kernel
- Gaussian process inference
  - Prediction & uncertainty quantification
- Observation noise
- What uncertainty are we quantifying?
- What can go wrong?
- Bayesian optimization

- Goals:
  - Learn the mechanism behind standard GPs to identify benefits and pitfalls (also in BayesOpt)
  - Learn the skills to be responsible users of standard GPs (transferable to other ML/AI methods)

Roadmap

- Bayesian modeling and inference
- Gaussian process model
  - Popular version using a squared exponential kernel
- Gaussian process inference
  - Prediction & uncertainty quantification
- Observation noise
- What uncertainty are we quantifying?
- What can go wrong?
- Bayesian optimization

- Goals:
  - Learn the mechanism behind standard GPs to identify benefits and pitfalls (also in BayesOpt)
  - Learn the skills to be responsible users of standard GPs (transferable to other ML/AI methods)

Roadmap

- Bayesian modeling and inference
- Gaussian process model
  - Popular version using a squared exponential kernel
- Gaussian process inference
  - Prediction & uncertainty quantification
- Observation noise
- What uncertainty are we quantifying?
- What can go wrong?
- Bayesian optimization

- Goals:
  - Learn the mechanism behind standard GPs to identify benefits and pitfalls (also in BayesOpt)
  - Learn the skills to be responsible users of standard GPs (transferable to other ML/AI methods)

Roadmap

- Bayesian modeling and inference
- Gaussian process model
  - Popular version using a squared exponential kernel
- Gaussian process inference
  - Prediction & uncertainty quantification
- Observation noise
- What uncertainty are we quantifying?
- What can go wrong?
- Bayesian optimization

- Goals:
  - Learn the mechanism behind standard GPs to identify benefits and pitfalls (also in BayesOpt)
  - Learn the skills to be responsible users of standard GPs (transferable to other ML/AI methods)

A Bayesian approach

A Bayesian approach

- p(unknowns | data)

A Bayesian approach

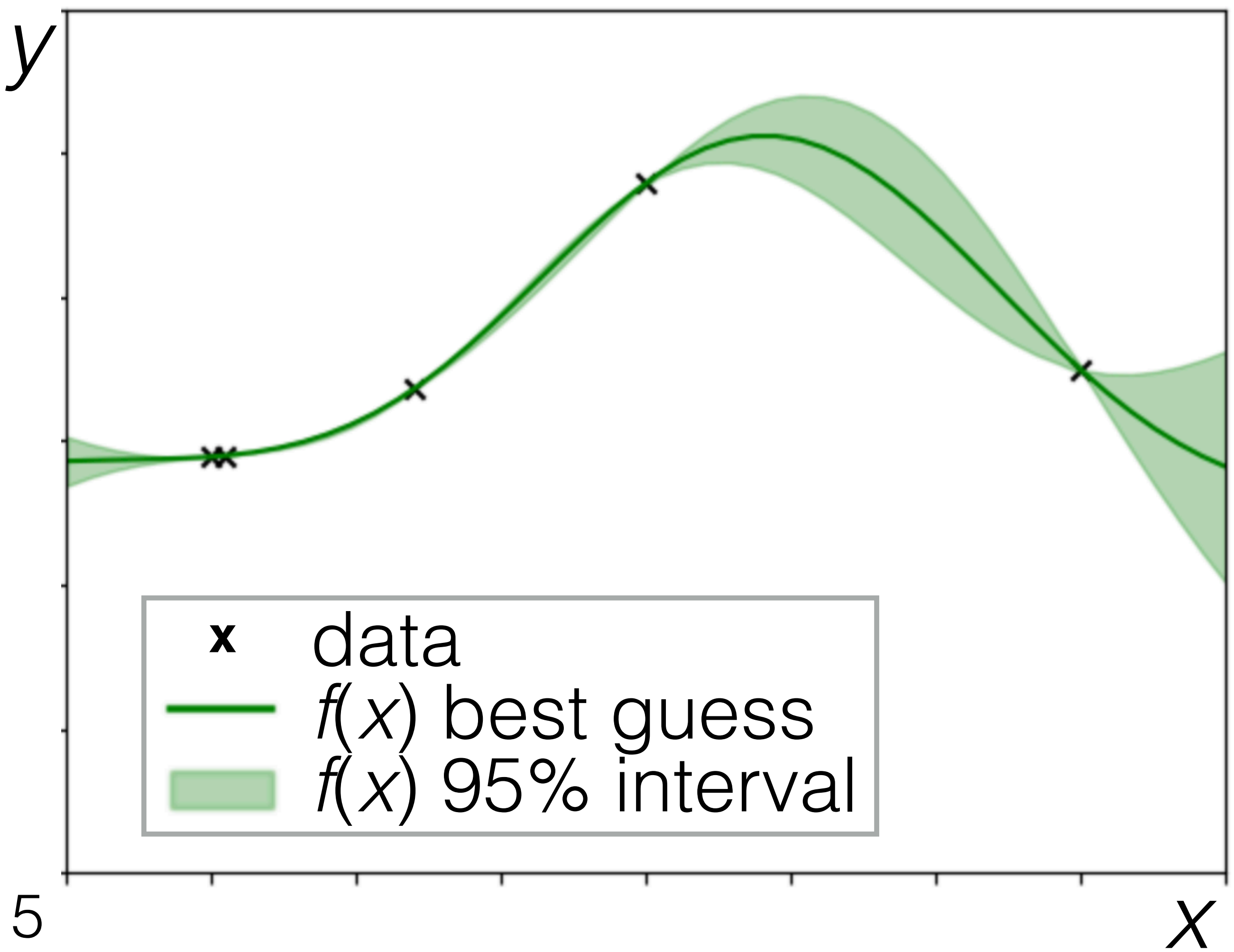
- p(unknowns | data)

Given the data we've seen, what do we know about the underlying function?

A Bayesian approach

- p(unknowns | data)

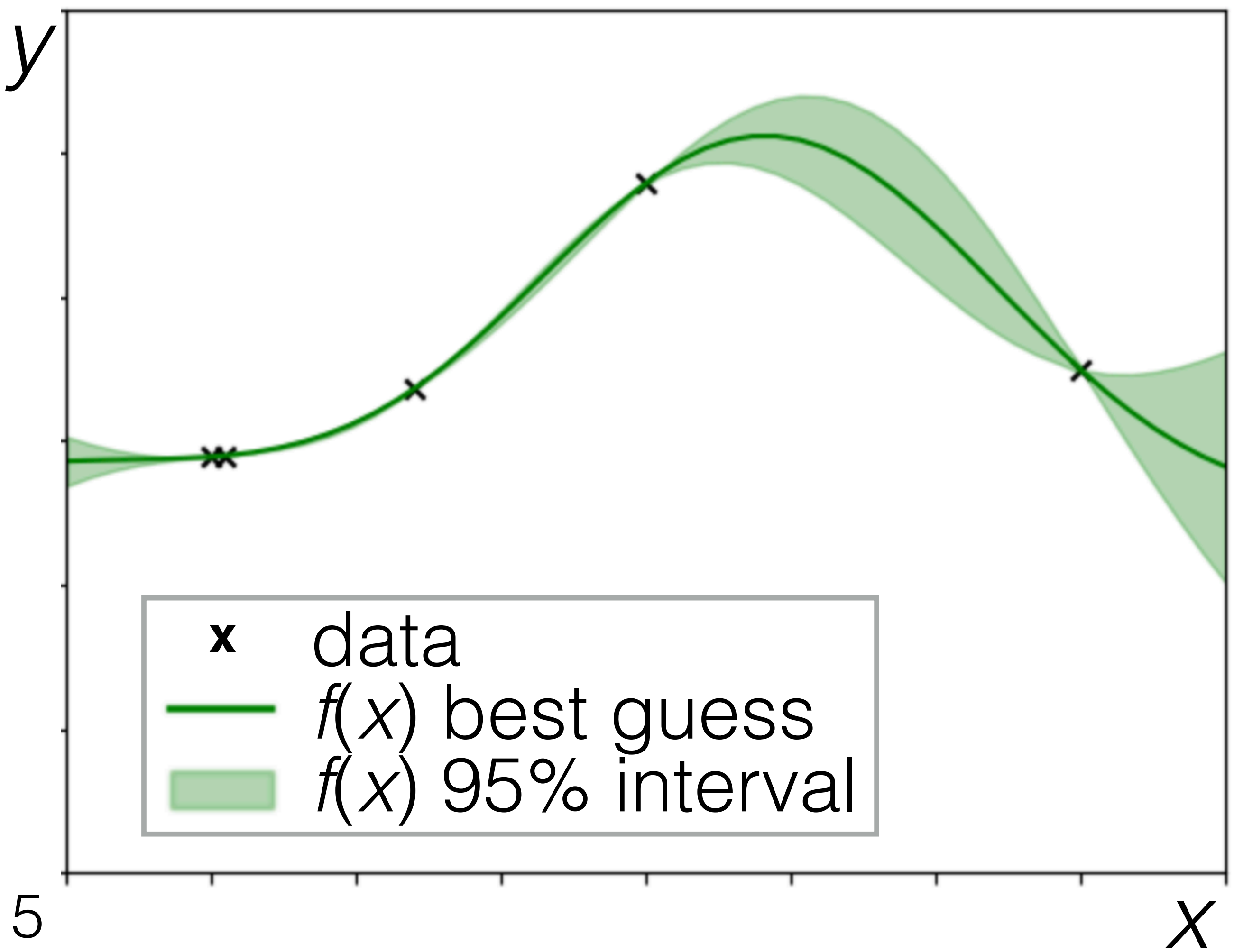
Given the data we've seen, what do we know about the underlying function?



A Bayesian approach

- p(unknowns | data) $\propto$ p(data | unknowns) p(unknowns)

Given the data we've seen, what do we know about the underlying function?

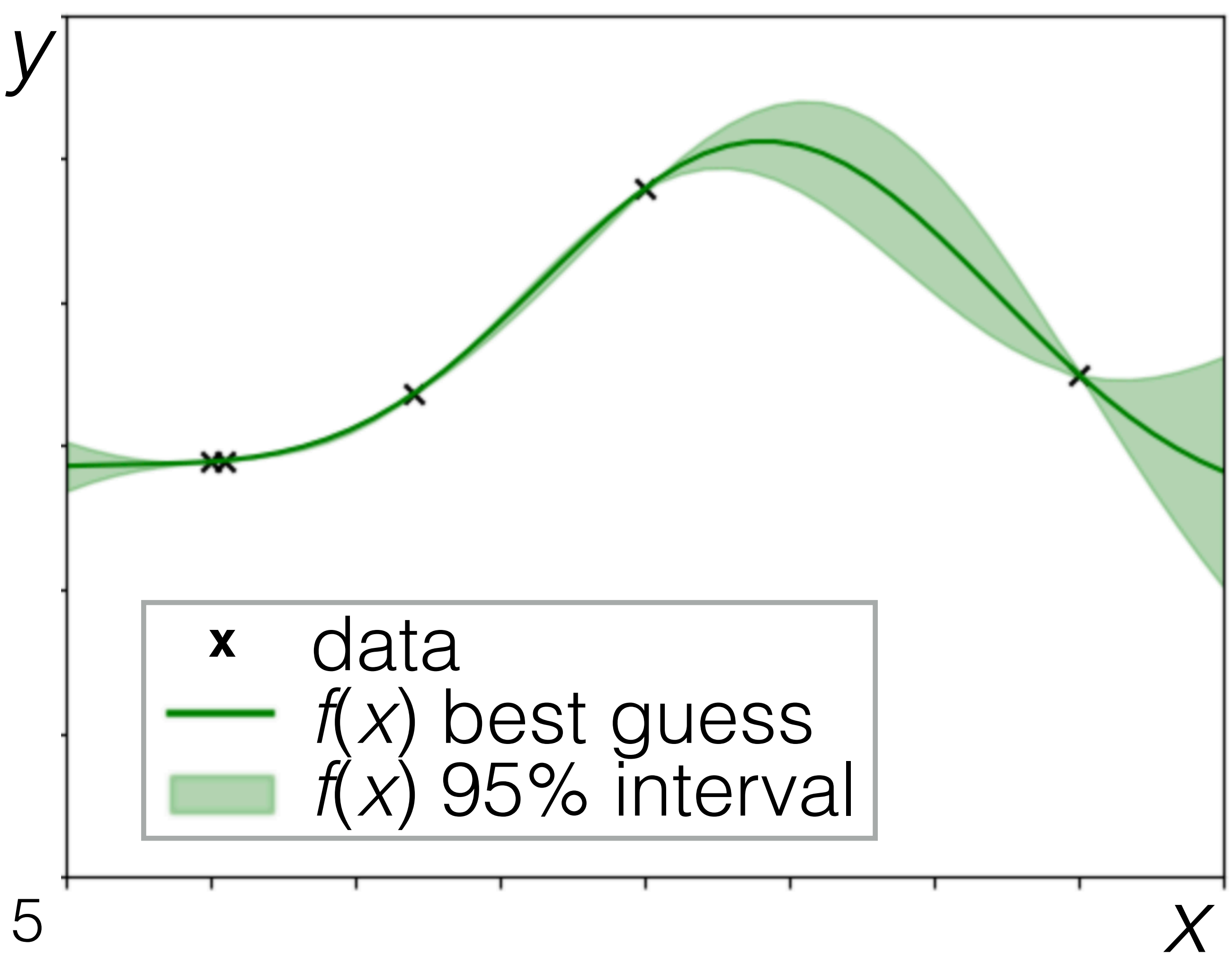


A Bayesian approach

- p(unknowns | data) $\propto$ p(data | unknowns) p(unknowns)

Given the data we've seen, what do we know about the underlying function?

A (statistical) model that can generate functions and data of interest

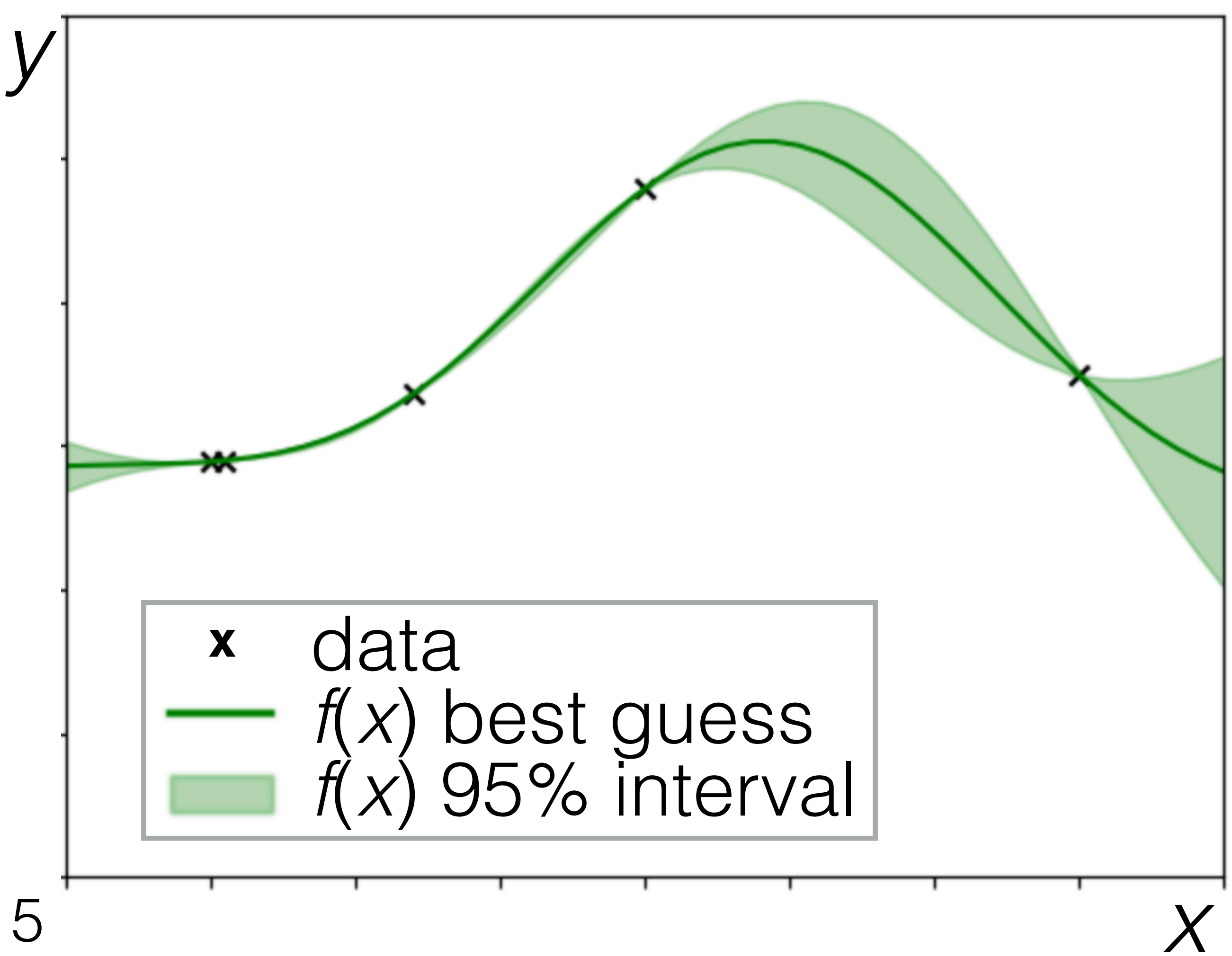


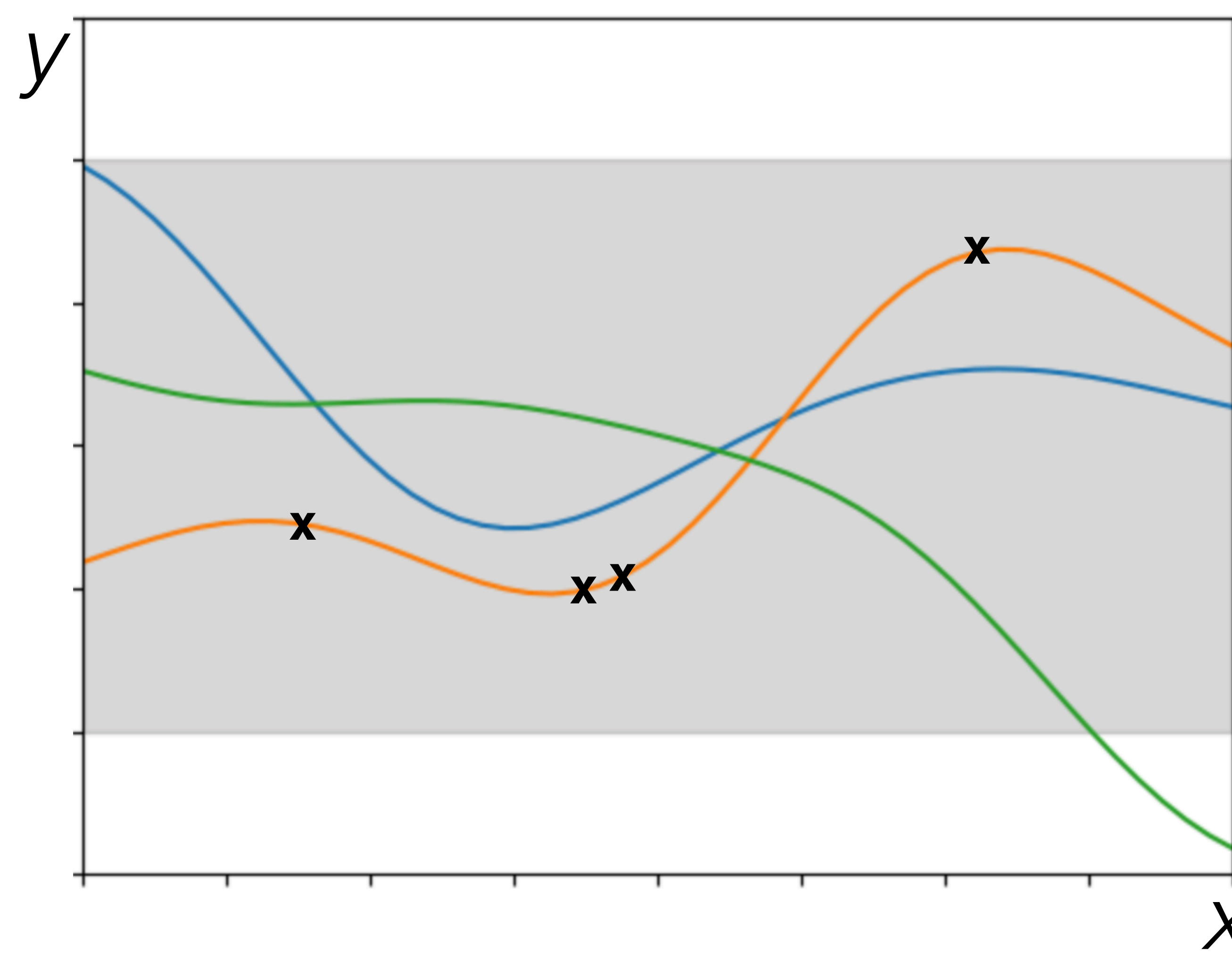
A Bayesian approach

- p(unknowns | data) ∝ p(data | unknowns) p(unknowns)

Given the data we've seen, what do we know about the underlying function?

A (statistical) model that can generate functions and data of interest





Univariate Gaussian distribution review

Univariate Gaussian distribution review

$$\mathcal{N}(y|\mu, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left[\frac{-(y-\mu)^2}{\sigma^2}\right]$$

$y \in (-\infty, \infty)$

Mean $\mu \in (-\infty, \infty)$

Standard deviation $\sigma \in (0, \infty)$

(Variance σ^2)

Univariate Gaussian distribution review

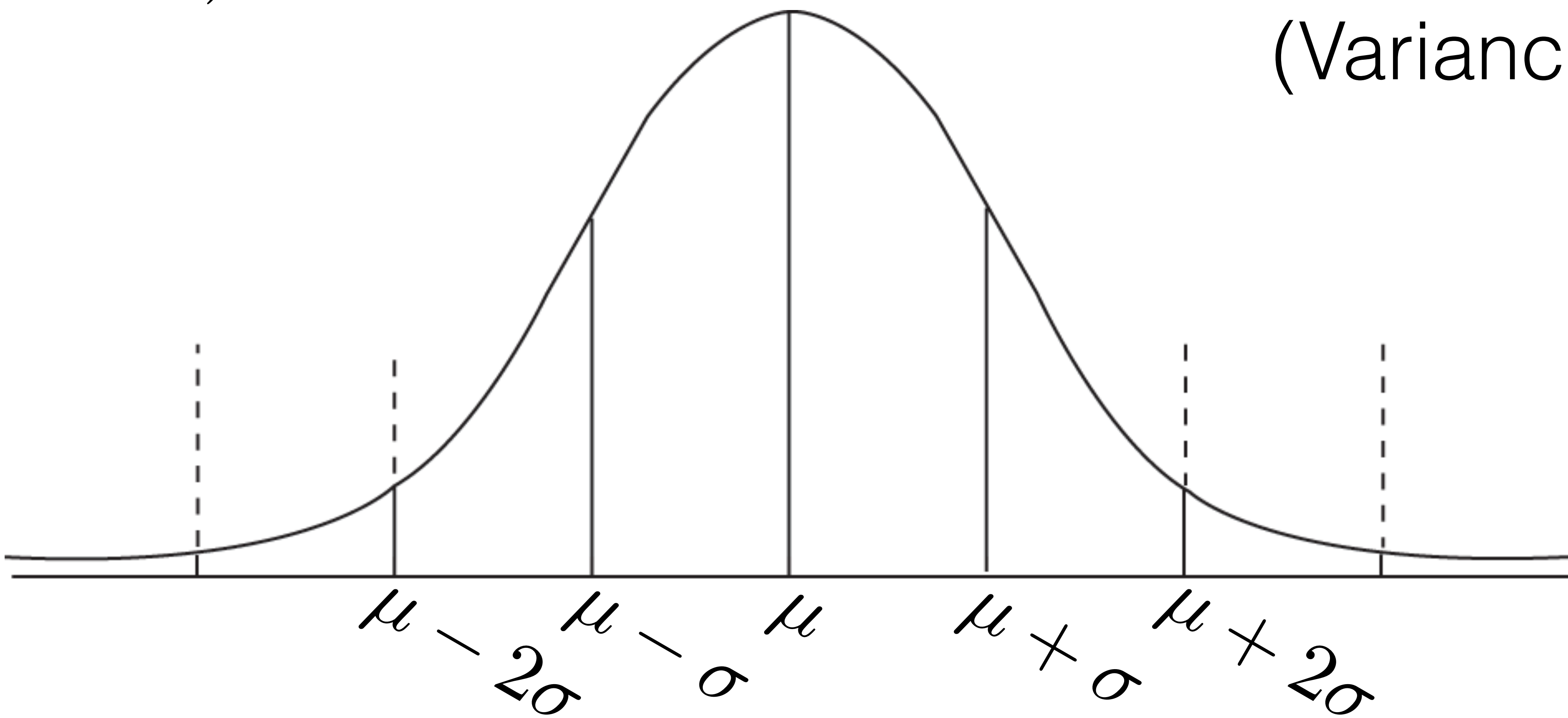
$$\mathcal{N}(y|\mu,\sigma^2) = (2\pi\sigma^2)^{-1/2}\exp\left[\frac{-(y-\mu)^2}{\sigma^2}\right]$$

$y \in (-\infty, \infty)$

Mean $\mu \in (-\infty, \infty)$

Standard deviation $\sigma \in (0, \infty)$

(Variance σ^2)



Univariate Gaussian distribution review

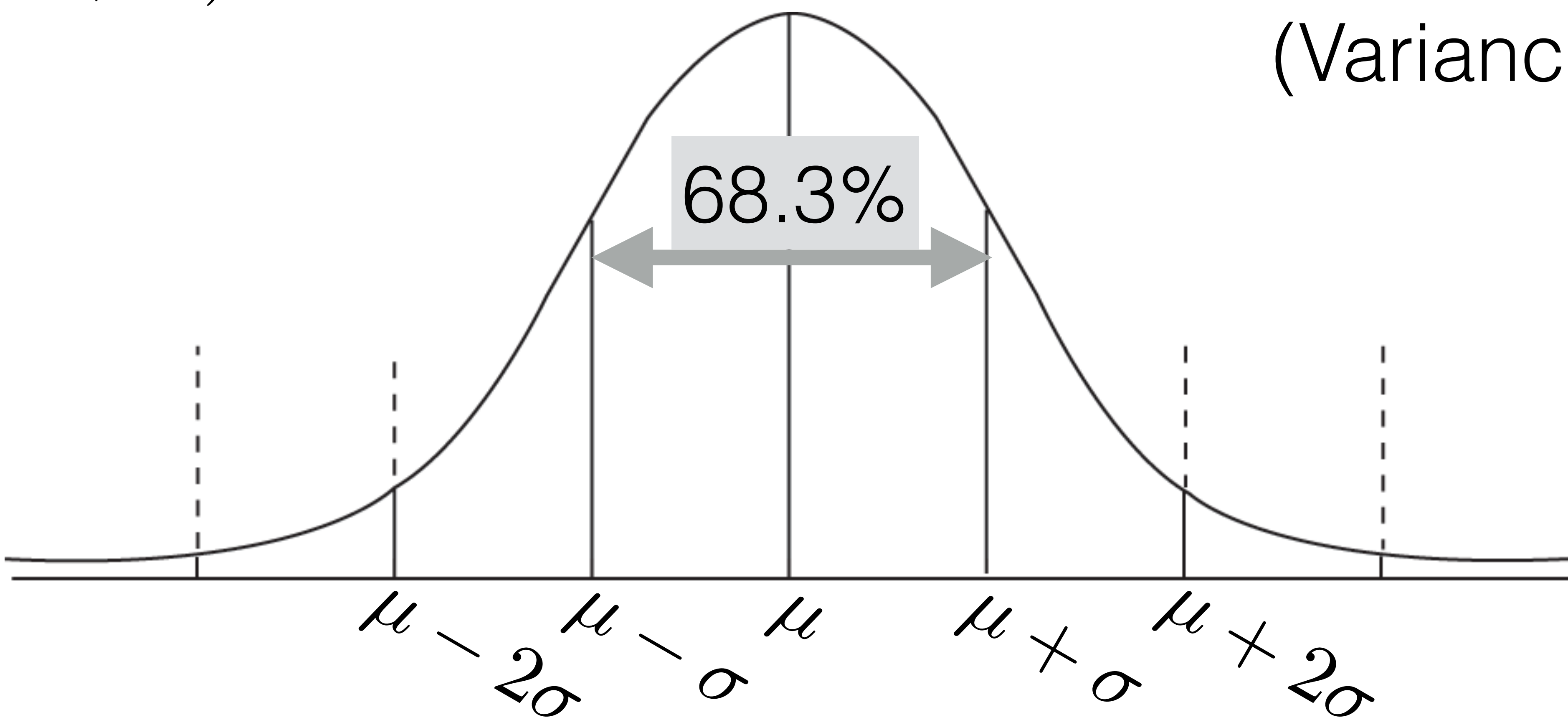
$$\mathcal{N}(y|\mu,\sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left[\frac{-(y-\mu)^2}{\sigma^2}\right]$$

$y \in (-\infty, \infty)$

Mean $\mu \in (-\infty, \infty)$

Standard deviation $\sigma \in (0, \infty)$

(Variance σ^2)



Univariate Gaussian distribution review

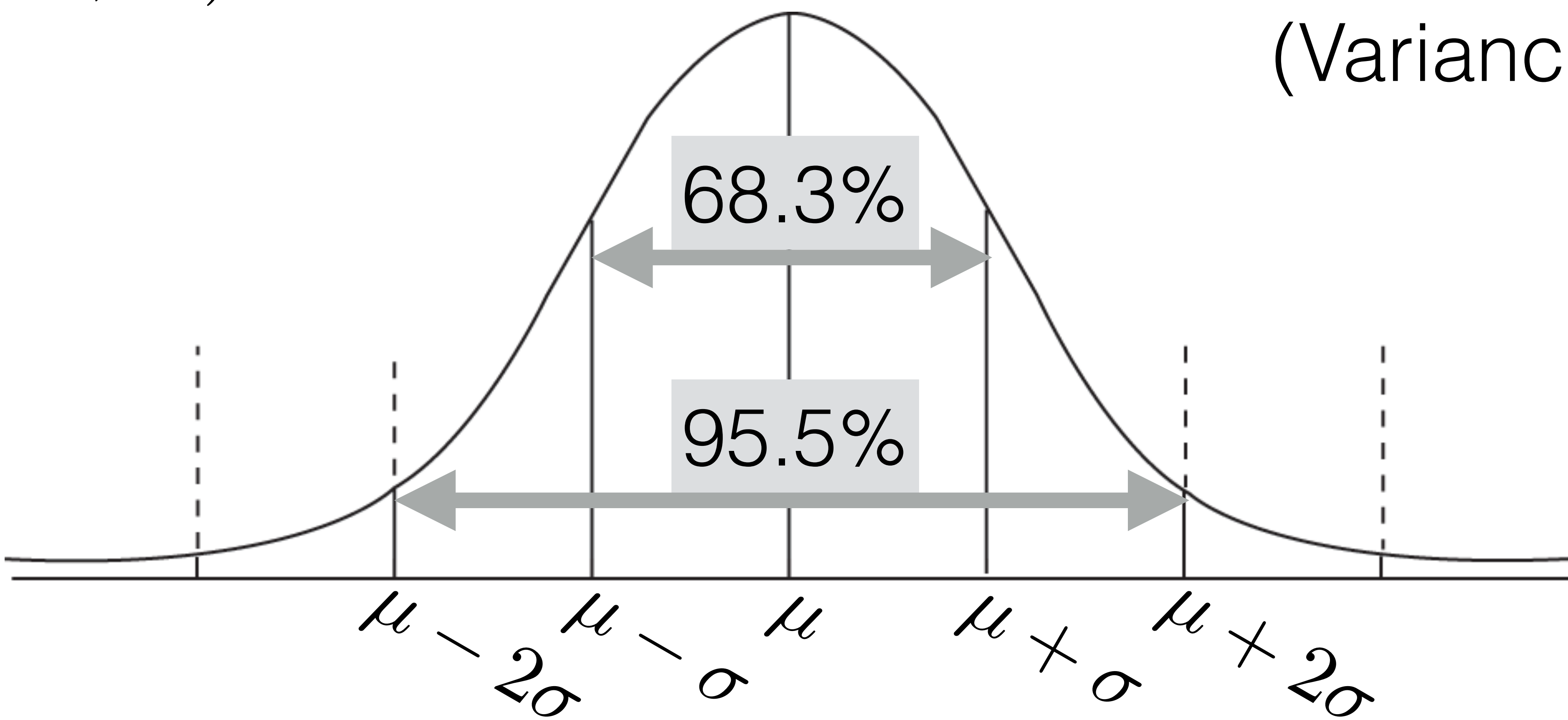
$$\mathcal{N}(y|\mu, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left[\frac{-(y-\mu)^2}{\sigma^2}\right]$$

$y \in (-\infty, \infty)$

Mean $\mu \in (-\infty, \infty)$

Standard deviation $\sigma \in (0, \infty)$

(Variance σ^2)



Univariate Gaussian distribution review

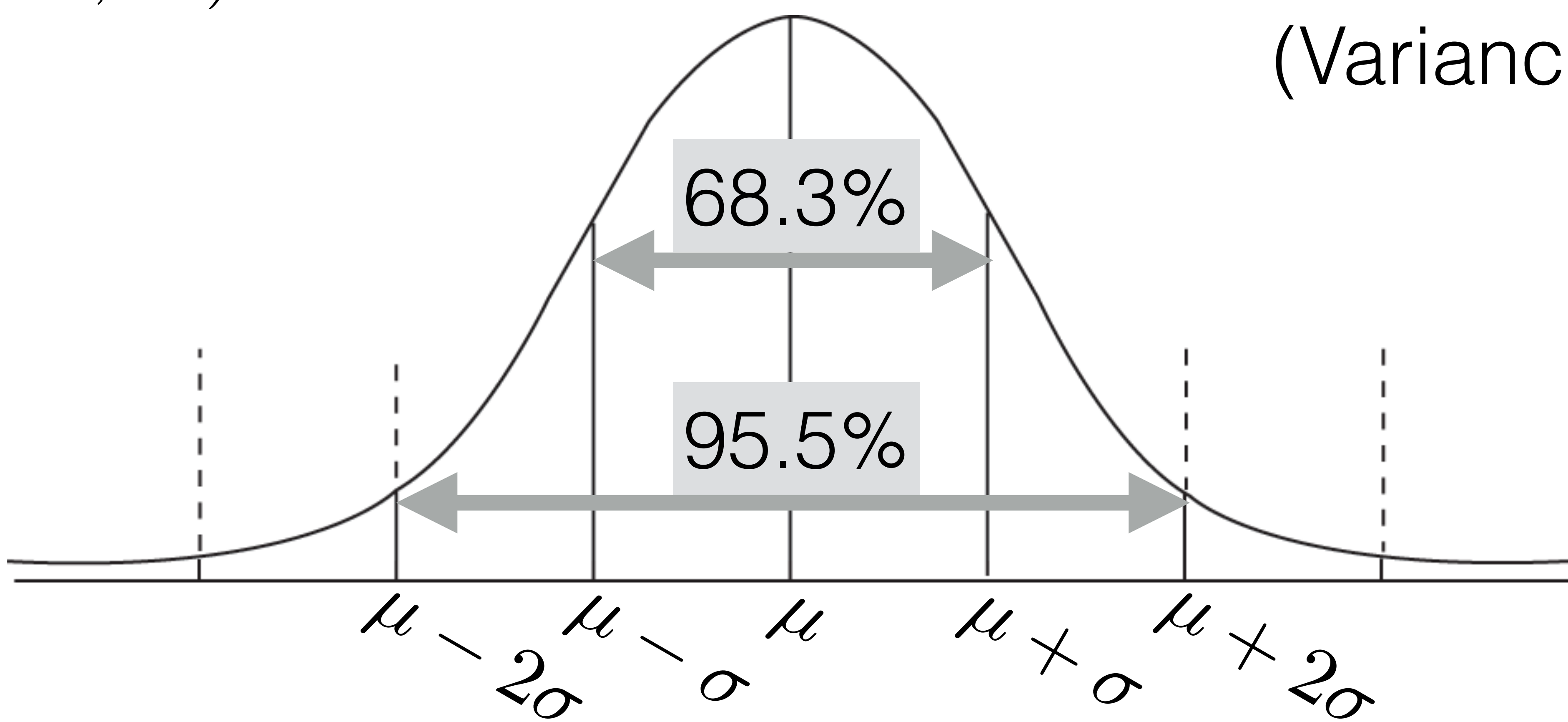
$$\mathcal{N}(y|\mu,\sigma^2) = (2\pi\sigma^2)^{-1/2}\exp\left[\frac{-(y-\mu)^2}{\sigma^2}\right]$$

$y \in (-\infty, \infty)$

Mean $\mu \in (-\infty, \infty)$

Standard deviation $\sigma \in (0, \infty)$

(Variance σ^2)



- About 95% of the mass falls within 2 standard deviations of the mean

Univariate Gaussian distribution review

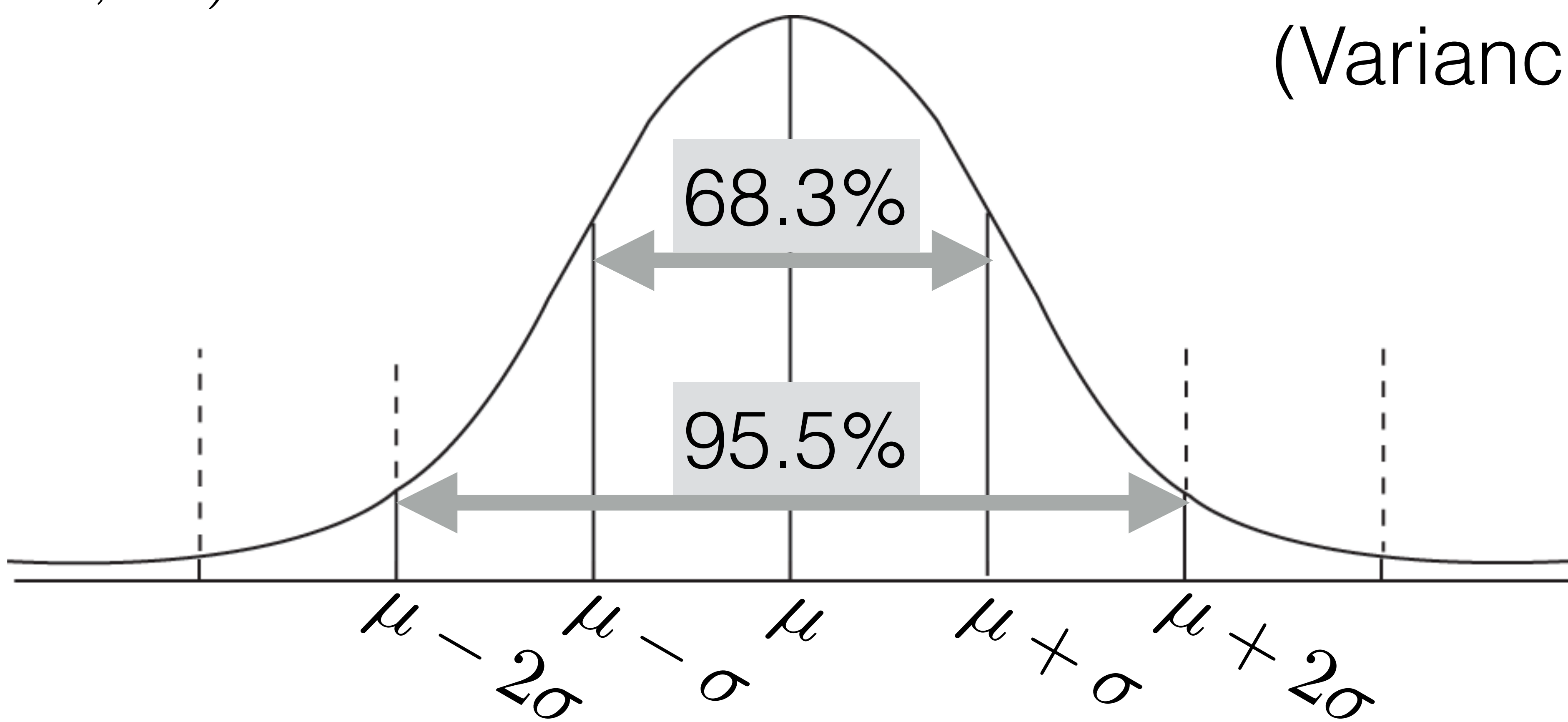
$$\mathcal{N}(y|\mu,\sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left[\frac{-(y-\mu)^2}{\sigma^2}\right]$$

$y \in (-\infty, \infty)$

Mean $\mu \in (-\infty, \infty)$

Standard deviation $\sigma \in (0, \infty)$

(Variance σ^2)



- About 95% of the mass falls within 2 standard deviations of the mean

[demo]

Univariate Gaussian distribution review

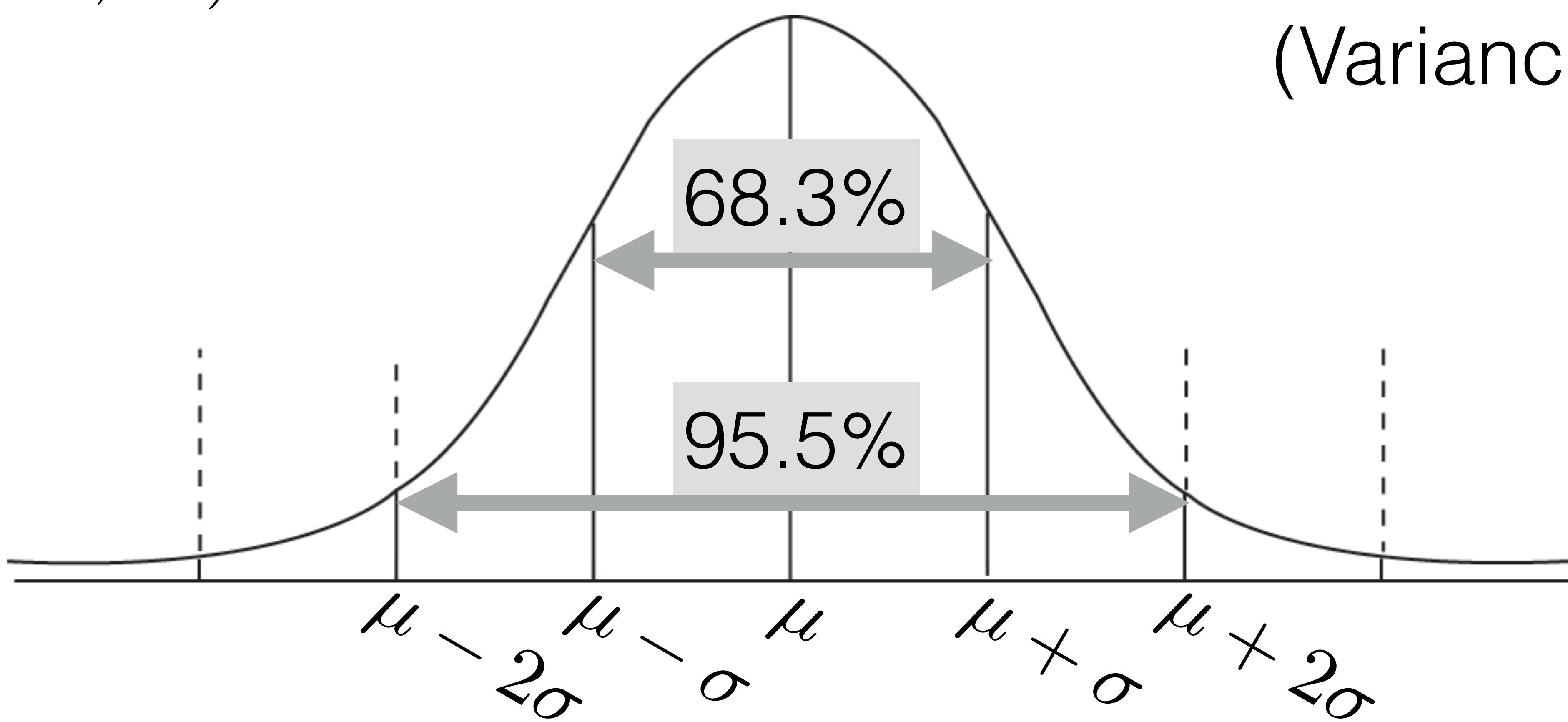
$$\mathcal{N}(y|\mu,\sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left[\frac{-(y-\mu)^2}{\sigma^2}\right]$$

$y \in (-\infty, \infty)$

Mean $\mu \in (-\infty, \infty)$

Standard deviation $\sigma \in (0, \infty)$

(Variance σ^2)



- About 95% of the mass falls within 2 standard deviations of the mean

[demo]

- If $Y \sim \mathcal{N}(0,1)$, then $Y + \mu \sim \mathcal{N}(\mu, 1)$

$$\sigma Y \sim \mathcal{N}(0, \sigma^2)$$

Multivariate Gaussian review

Multivariate Gaussian review

- General form: $\mathcal{N}(y|\mu, K)$ with
  - y & the mean μ are real-valued vectors with dimension *M*
  - Covariance matrix *K* is symmetric, positive semidefinite

Multivariate Gaussian review

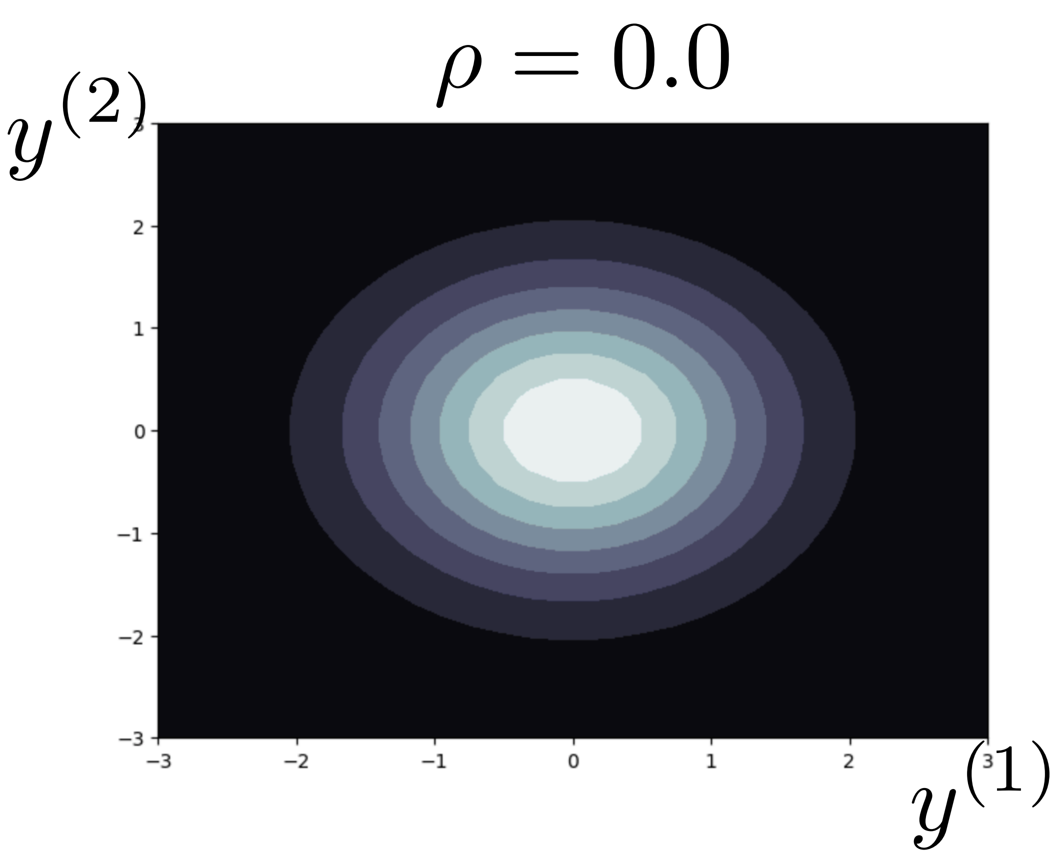
- General form: $\mathcal{N}(y|\mu, K)$ with
  - y & the mean μ are real-valued vectors with dimension *M*
  - Covariance matrix *K* is symmetric, positive semidefinite
- Special case: *M* = 2, so $y = [y^{(1)}, y^{(2)}]^\top$ & $\mu = [\mu^{(1)}, \mu^{(2)}]^\top$

Multivariate Gaussian review

- General form: $\mathcal{N}(y|\mu, K)$ with
  - y & the mean μ are real-valued vectors with dimension *M*
  - Covariance matrix *K* is symmetric, positive semidefinite
- Special case: *M* = 2, so $y = [y^{(1)}, y^{(2)}]^\top$ & $\mu = [\mu^{(1)}, \mu^{(2)}]^\top$
  - Let's also assume $K = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ ← correlation

Multivariate Gaussian review

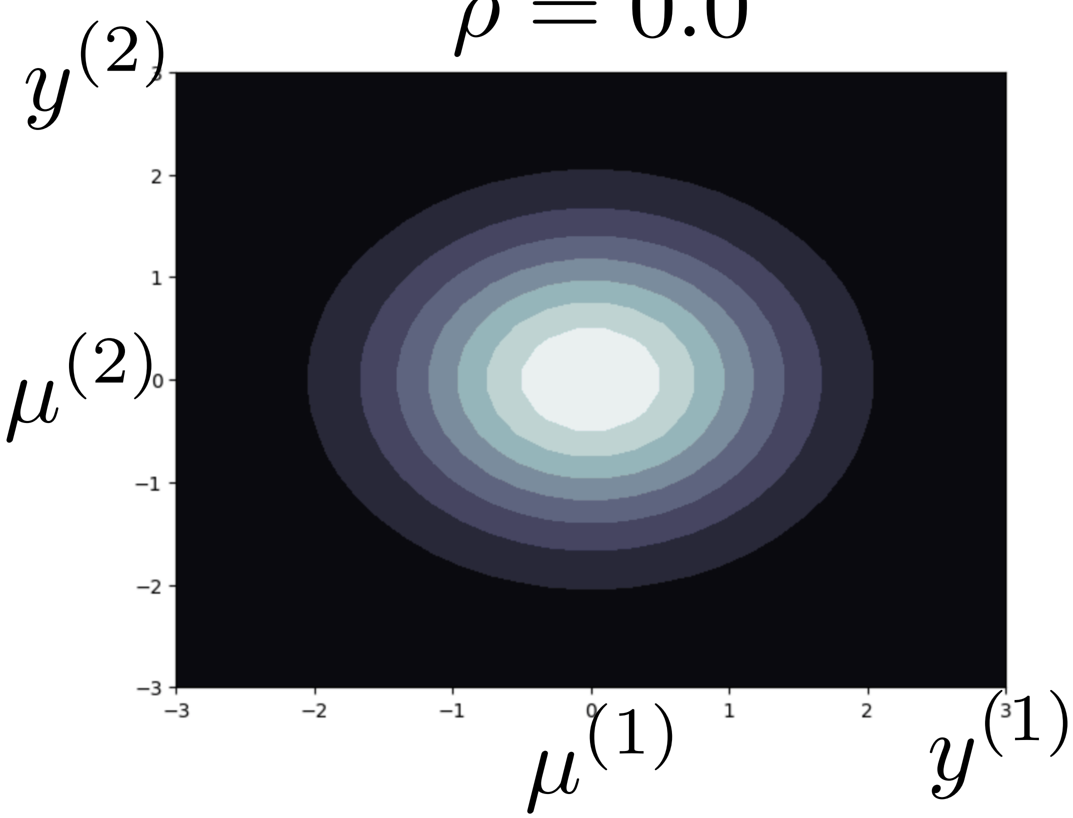
- General form: $\mathcal{N}(y|\mu, K)$ with
  - y & the mean μ are real-valued vectors with dimension M
  - Covariance matrix K is symmetric, positive semidefinite
- Special case: $M = 2$, so $y = [y^{(1)}, y^{(2)}]^\top$ & $\mu = [\mu^{(1)}, \mu^{(2)}]^\top$
  - Let's also assume $K = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ ← correlation



Multivariate Gaussian review

- General form: $\mathcal{N}(y|\mu, K)$ with
  - y & the mean μ are real-valued vectors with dimension *M*
  - Covariance matrix *K* is symmetric, positive semidefinite
- Special case: *M* = 2, so $y = [y^{(1)}, y^{(2)}]^\top$ & $\mu = [\mu^{(1)}, \mu^{(2)}]^\top$
  - Let's also assume $K = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ ← correlation

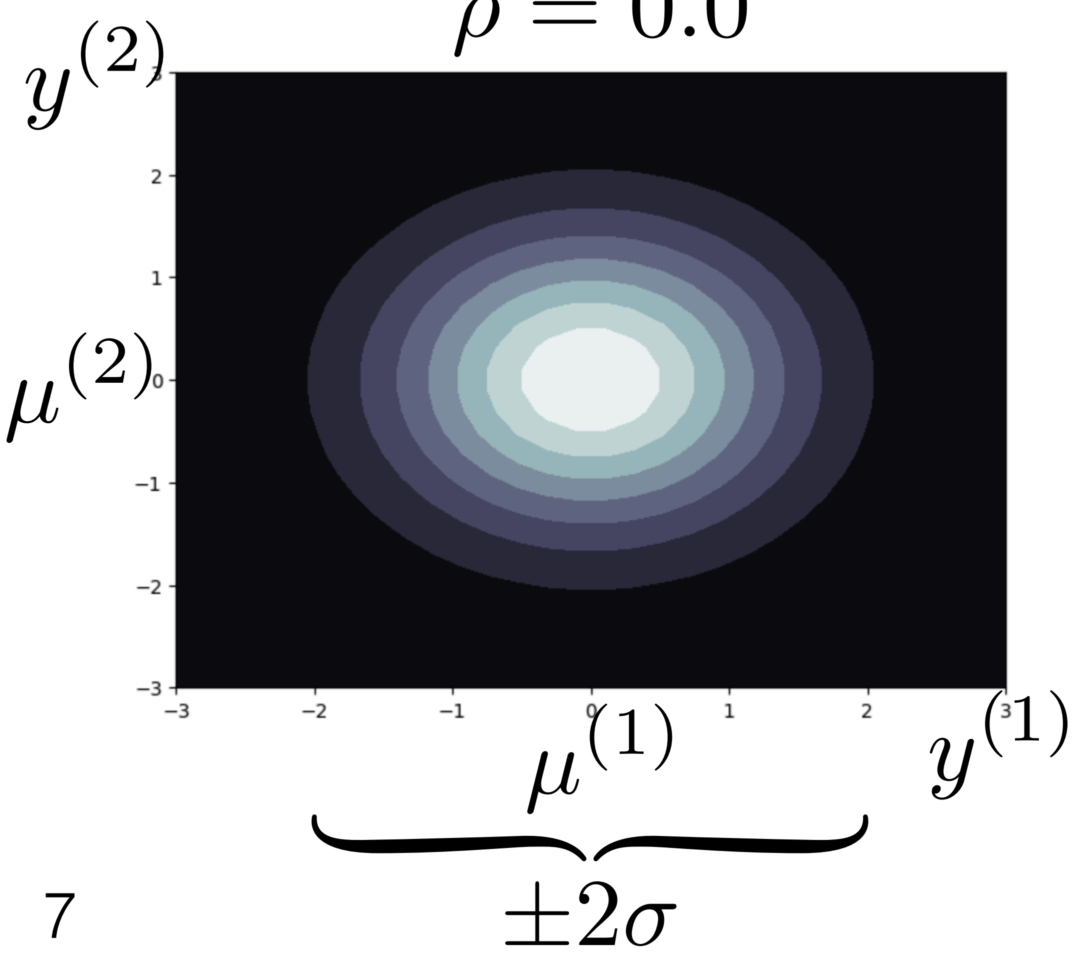
$\rho = 0.0$



Multivariate Gaussian review

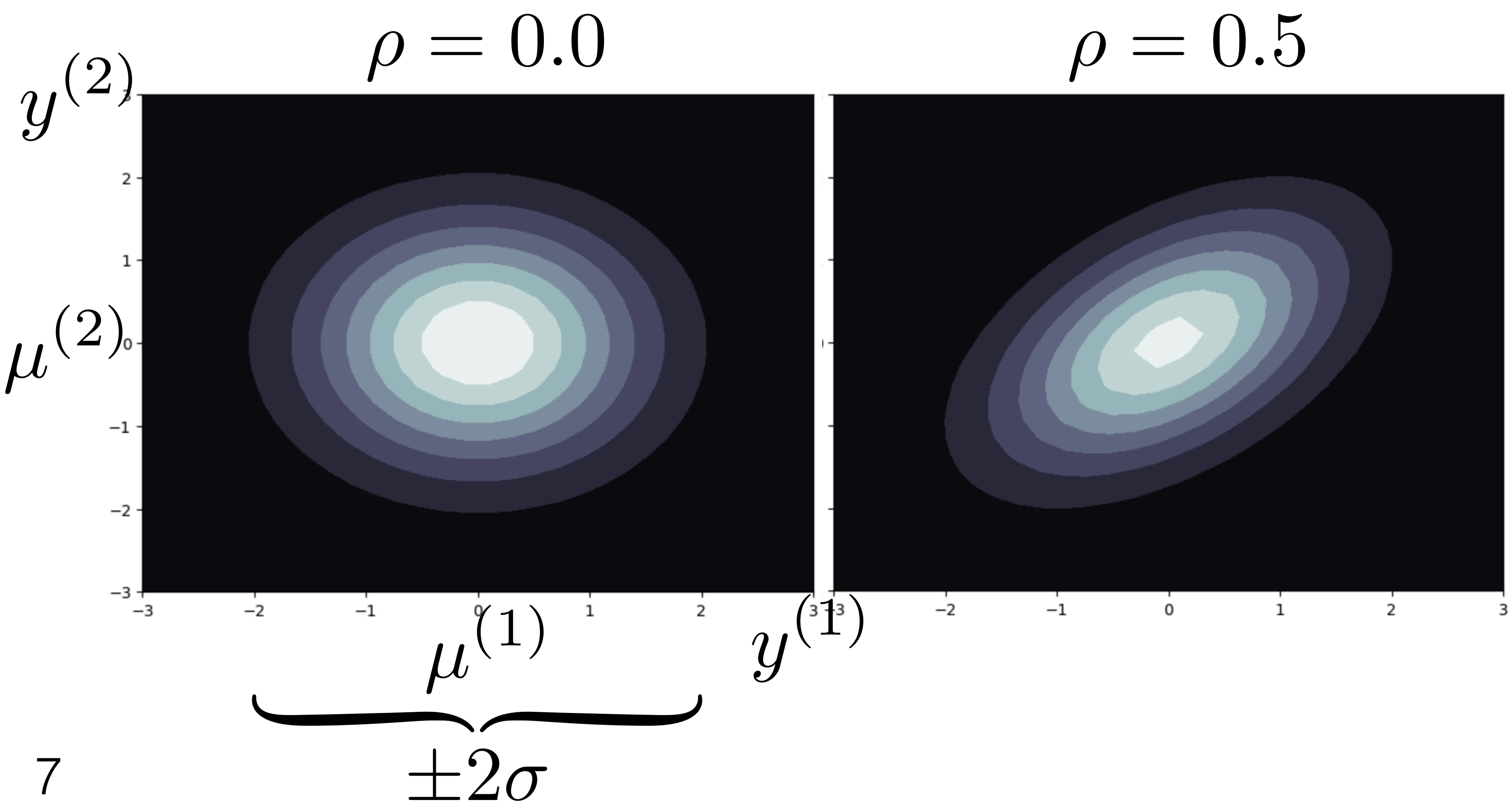
- General form: $\mathcal{N}(y|\mu, K)$ with
  - y & the mean μ are real-valued vectors with dimension *M*
  - Covariance matrix *K* is symmetric, positive semidefinite
- Special case: *M* = 2, so $y = [y^{(1)}, y^{(2)}]^\top$ & $\mu = [\mu^{(1)}, \mu^{(2)}]^\top$
  - Let's also assume $K = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ ← correlation

$\rho = 0.0$



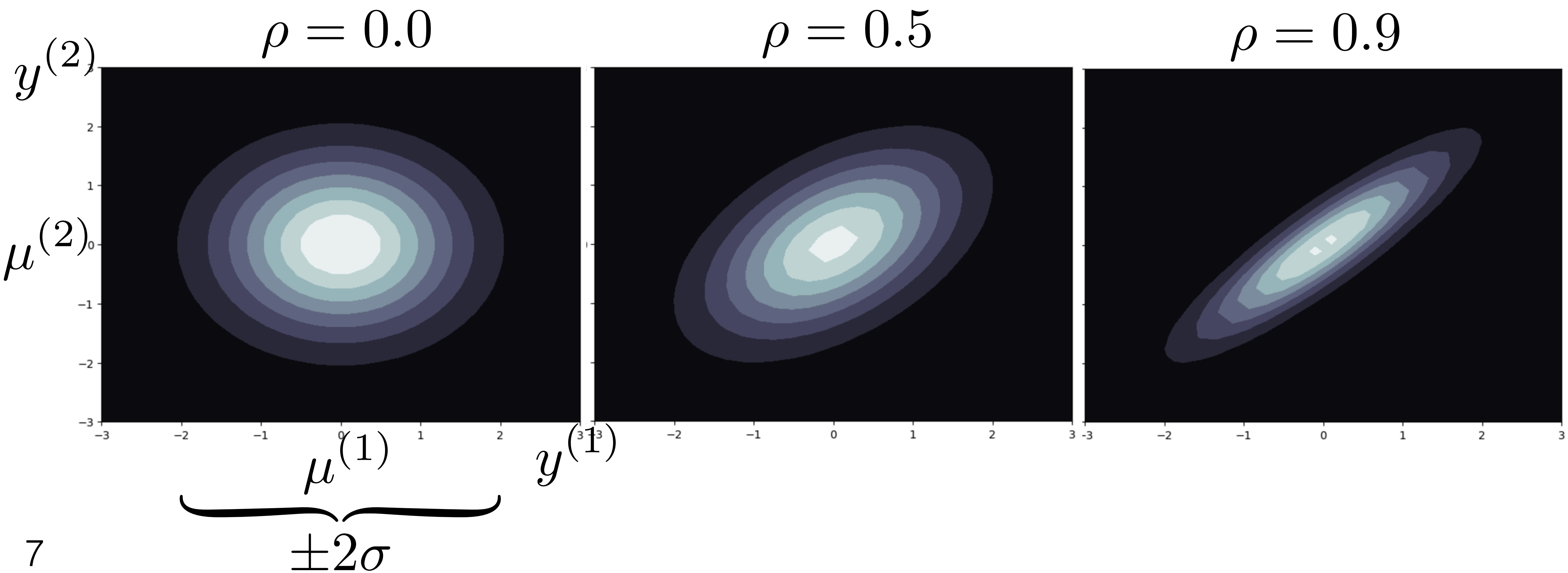
Multivariate Gaussian review

- General form: $\mathcal{N}(y|\mu, K)$ with
  - y & the mean μ are real-valued vectors with dimension *M*
  - Covariance matrix *K* is symmetric, positive semidefinite
- Special case: *M* = 2, so $y = [y^{(1)}, y^{(2)}]^\top$ & $\mu = [\mu^{(1)}, \mu^{(2)}]^\top$
  - Let's also assume $K = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ ← correlation



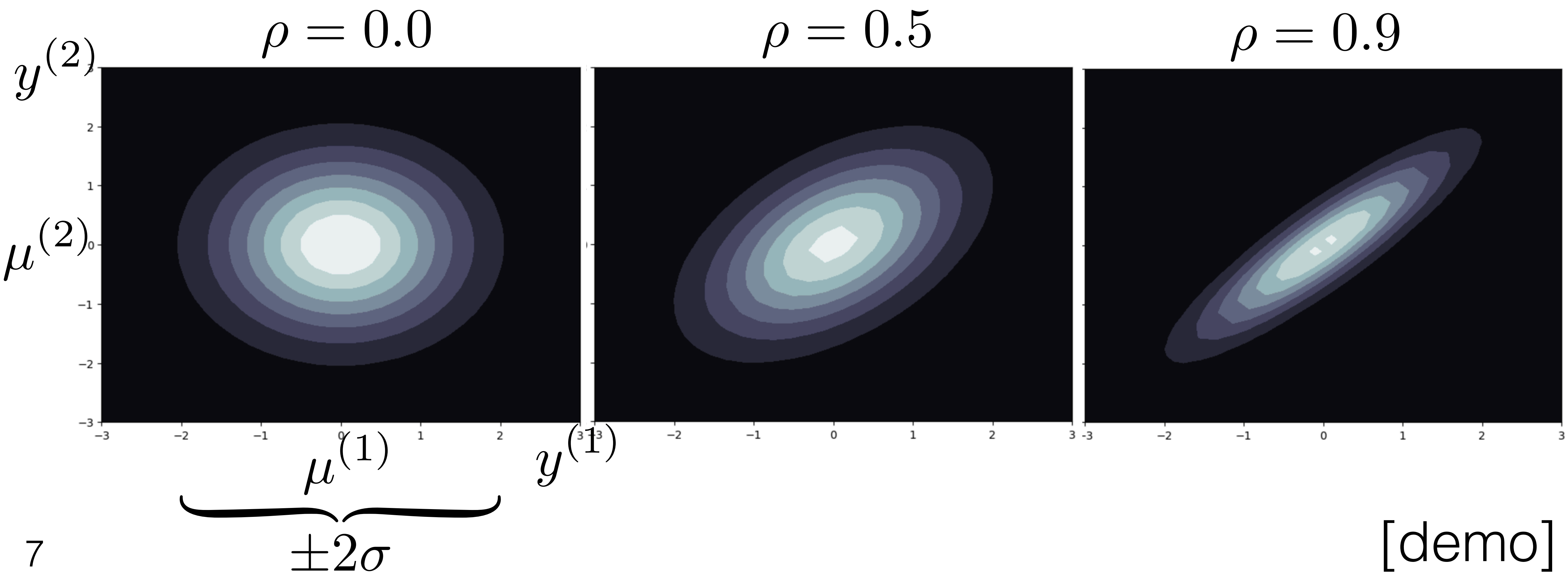
Multivariate Gaussian review

- General form: $\mathcal{N}(y|\mu, K)$ with
  - y & the mean μ are real-valued vectors with dimension *M*
  - Covariance matrix *K* is symmetric, positive semidefinite
- Special case: *M* = 2, so $y = [y^{(1)}, y^{(2)}]^\top$ & $\mu = [\mu^{(1)}, \mu^{(2)}]^\top$
  - Let's also assume $K = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ ← correlation



Multivariate Gaussian review

- General form: $\mathcal{N}(y|\mu, K)$ with
  - y & the mean μ are real-valued vectors with dimension *M*
  - Covariance matrix *K* is symmetric, positive semidefinite
- Special case: *M* = 2, so $y = [y^{(1)}, y^{(2)}]^\top$ & $\mu = [\mu^{(1)}, \mu^{(2)}]^\top$
  - Let's also assume $K = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ ← correlation



[demo]

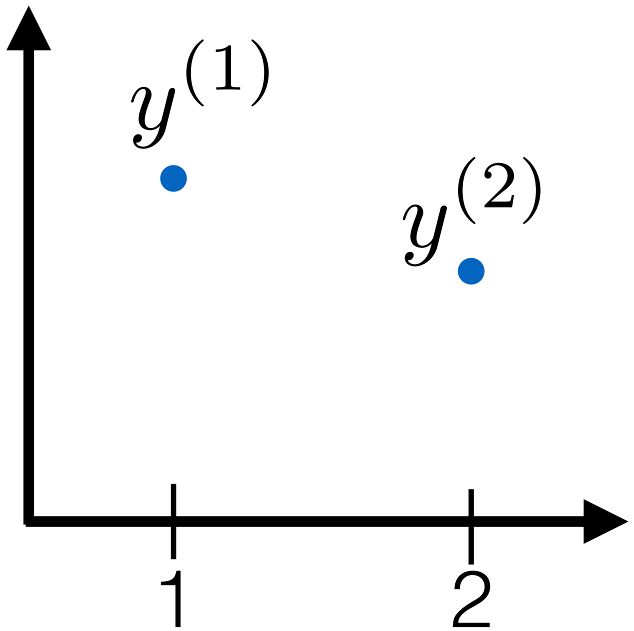
Multivariate Gaussian using locations

Multivariate Gaussian using locations

- *M*=2 dimensions: $y = [y_1, y_2]^\top \sim \mathcal{N}(\mu, K)$
  - With $\mu = [\mu_1, \mu_2]^\top$ and $K = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ ← correlation

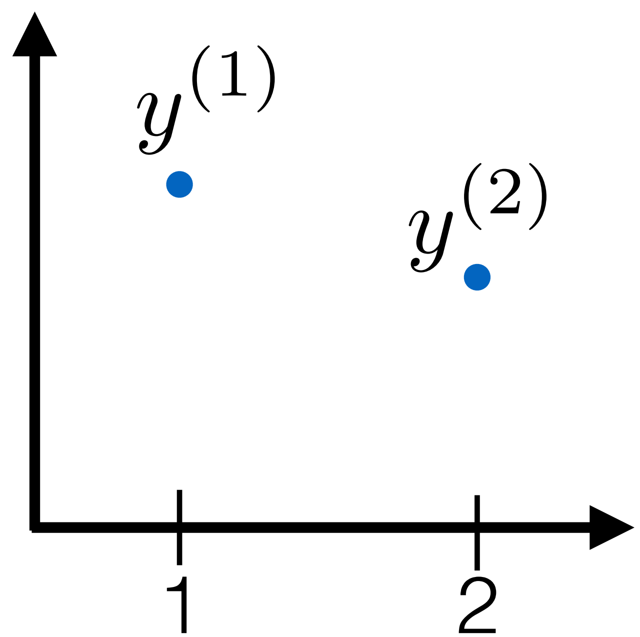
Multivariate Gaussian using locations

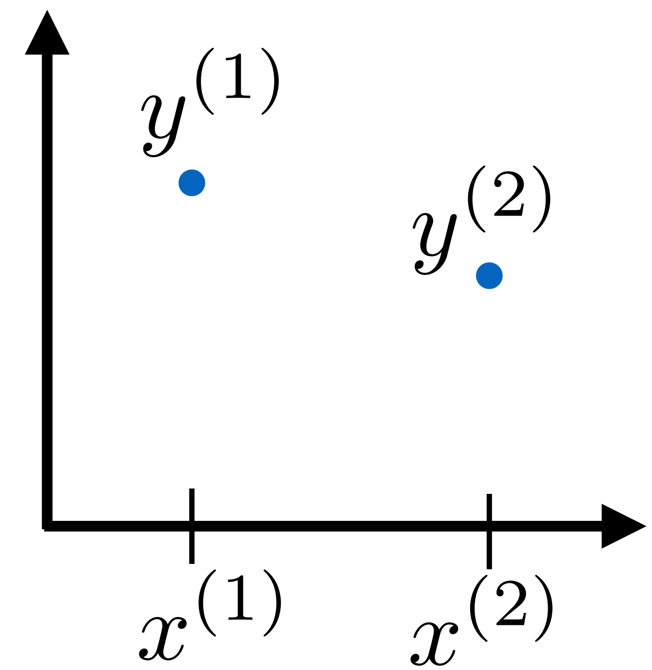
- *M*=2 dimensions: $y = [y_1, y_2]^\top \sim \mathcal{N}(\mu, K)$
  - With $\mu = [\mu_1, \mu_2]^\top$ and $K = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ ← correlation



Multivariate Gaussian using locations

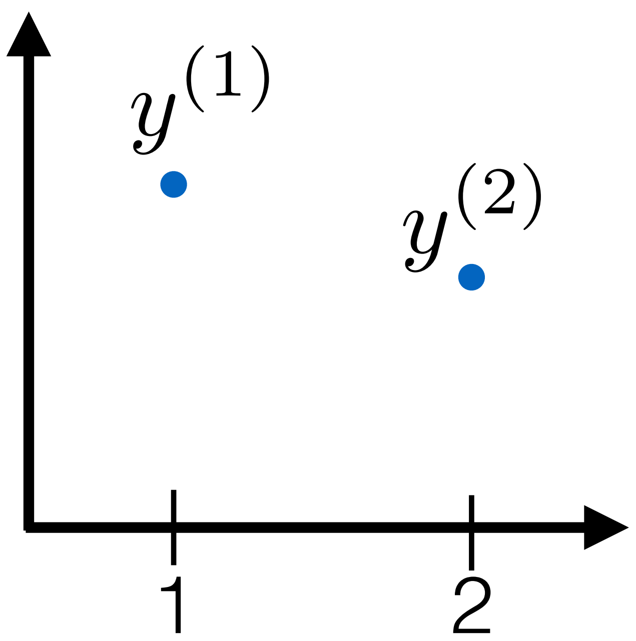
- M=2 dimensions: $y = [y_1, y_2]^\top \sim \mathcal{N}(\mu, K)$
  - With $\mu = [\mu_1, \mu_2]^\top$ and $K = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ ← correlation

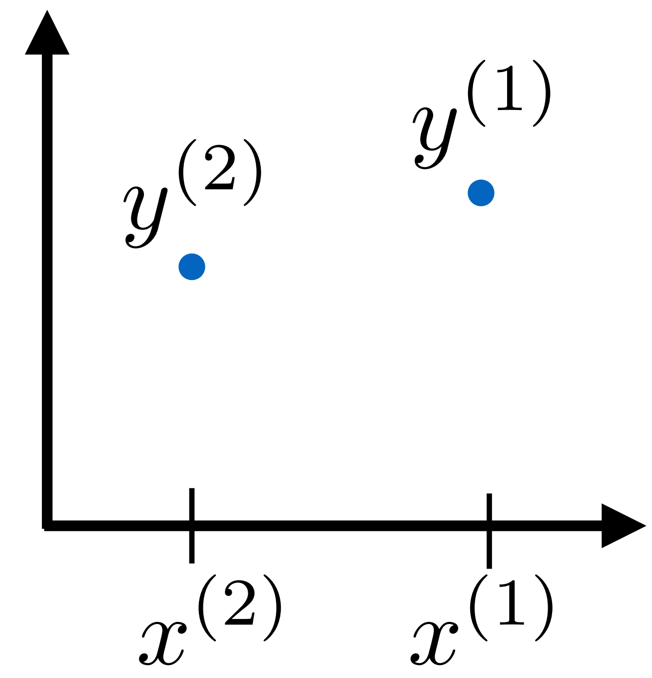




Multivariate Gaussian using locations

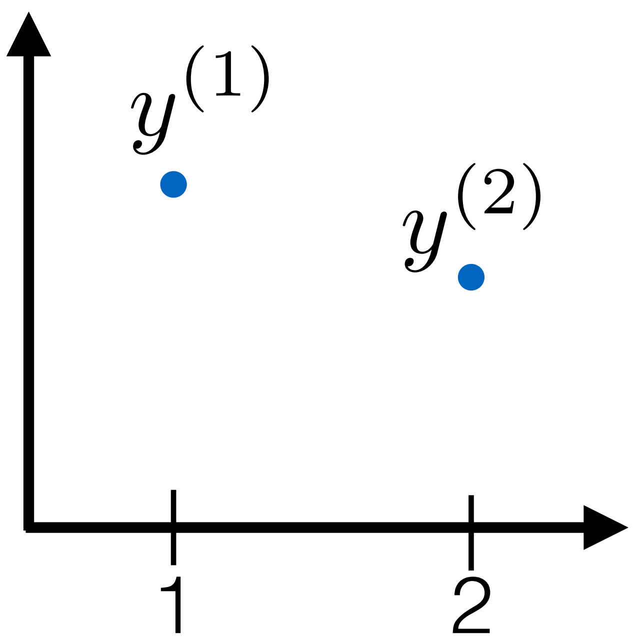
- *M*=2 dimensions: $y = [y_1, y_2]^\top \sim \mathcal{N}(\mu, K)$
  - With $\mu = [\mu_1, \mu_2]^\top$ and $K = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ ← correlation

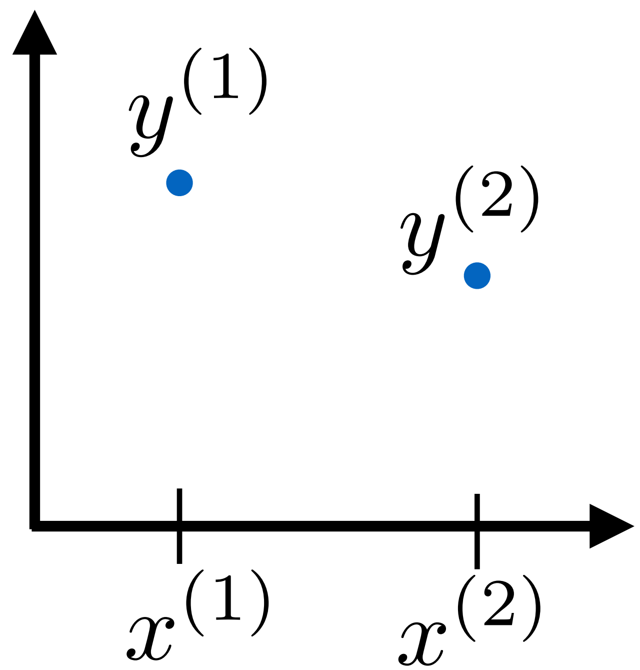




Multivariate Gaussian using locations

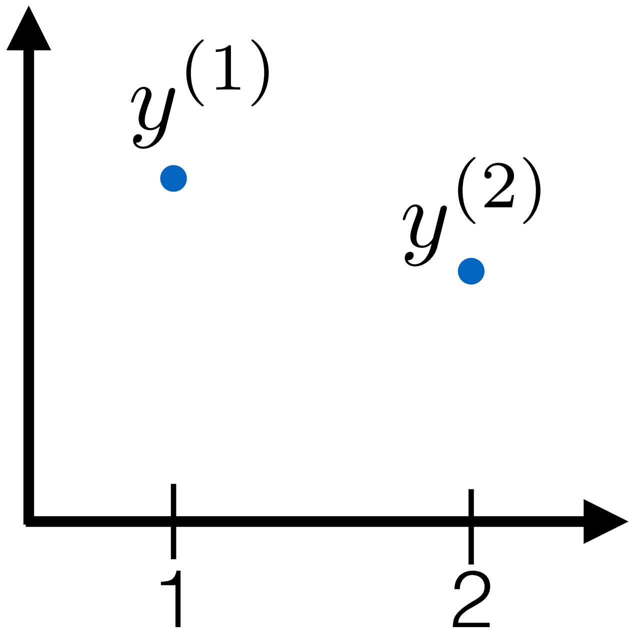
- M=2 dimensions: $y = [y_1, y_2]^\top \sim \mathcal{N}(\mu, K)$
  - With $\mu = [\mu_1, \mu_2]^\top$ and $K = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ ← correlation

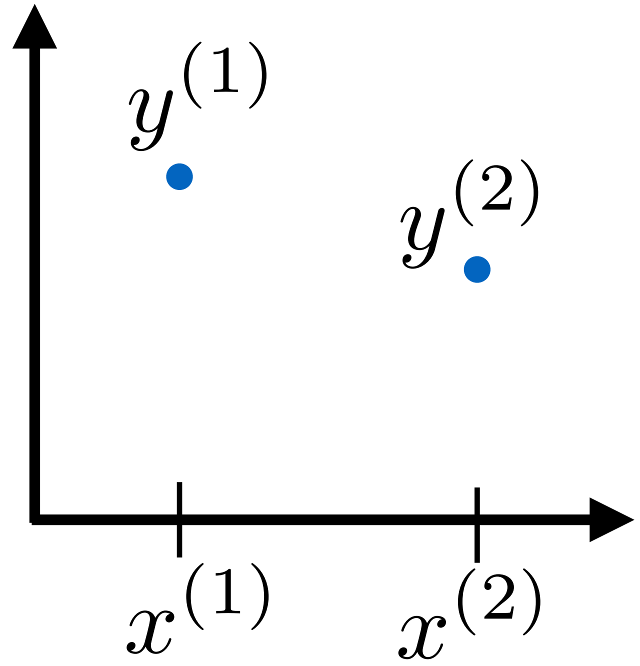




Multivariate Gaussian using locations

- *M*=2 dimensions: $y = [y_1, y_2]^\top \sim \mathcal{N}(\mu, K)$
  - With $\mu = [\mu_1, \mu_2]^\top$ and $K = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ ← correlation

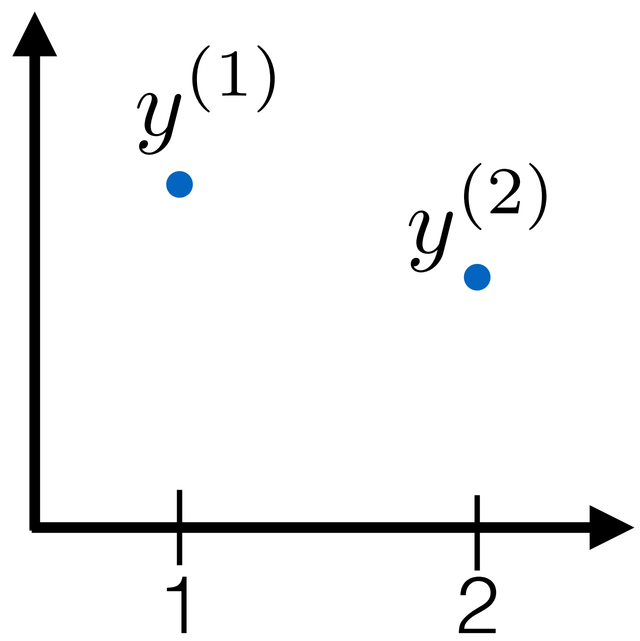


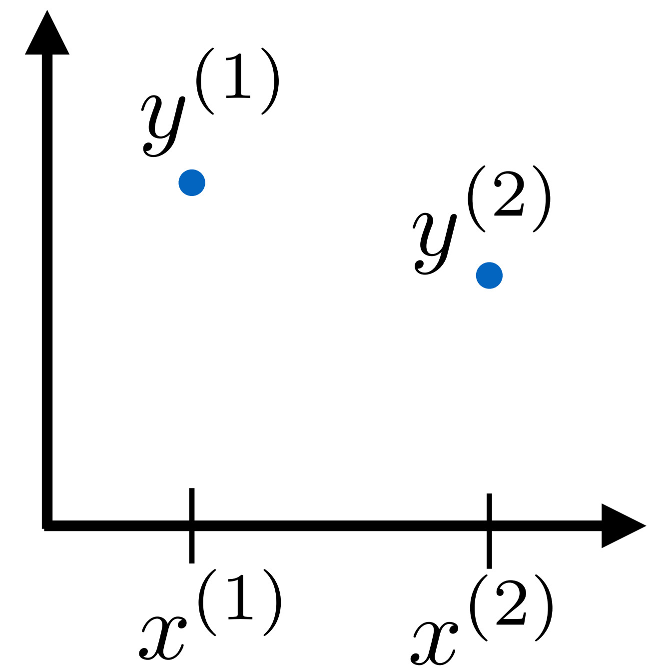


- What if we let the correlation depend on the *x*'s?

Multivariate Gaussian using locations

- M=2 dimensions: $y = [y_1, y_2]^\top \sim \mathcal{N}(\mu, K)$
  - With $\mu = [\mu_1, \mu_2]^\top$ and $K = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ ← correlation

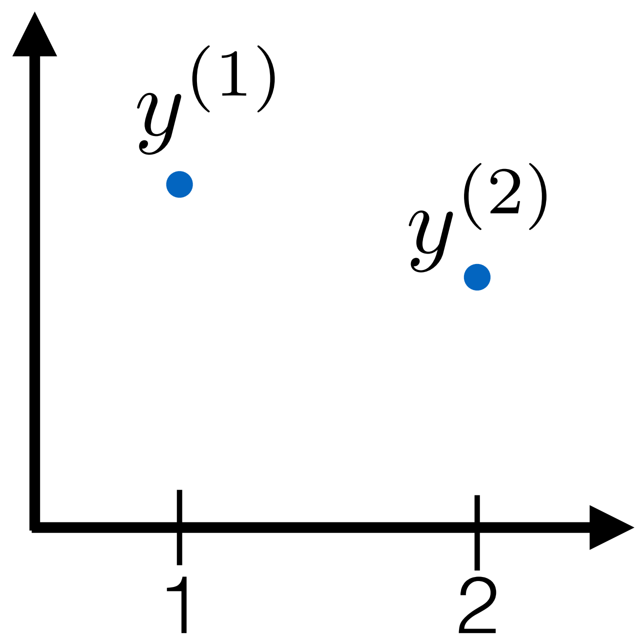


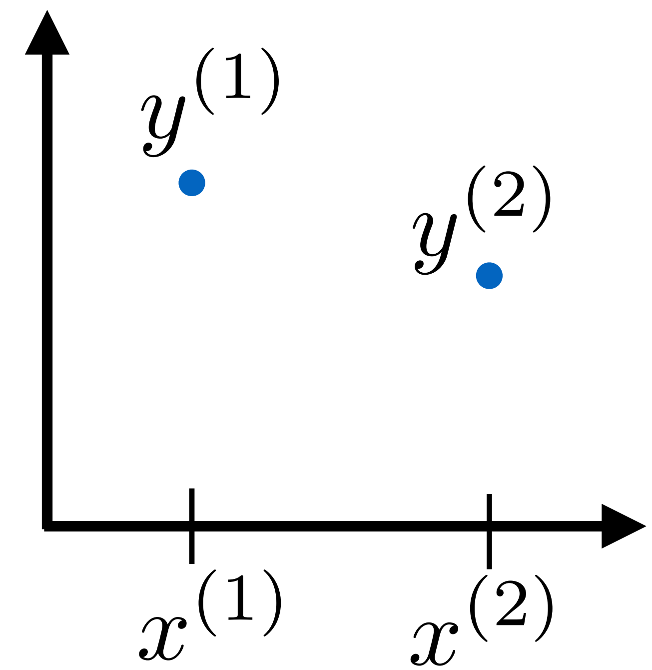


- What if we let the correlation depend on the *x*'s?
  - Let $\rho = \rho(|x^{(1)} - x^{(2)}|)$

Multivariate Gaussian using locations

- *M*=2 dimensions: $y = [y_1, y_2]^\top \sim \mathcal{N}(\mu, K)$
  - With $\mu = [\mu_1, \mu_2]^\top$ and $K = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ ← correlation

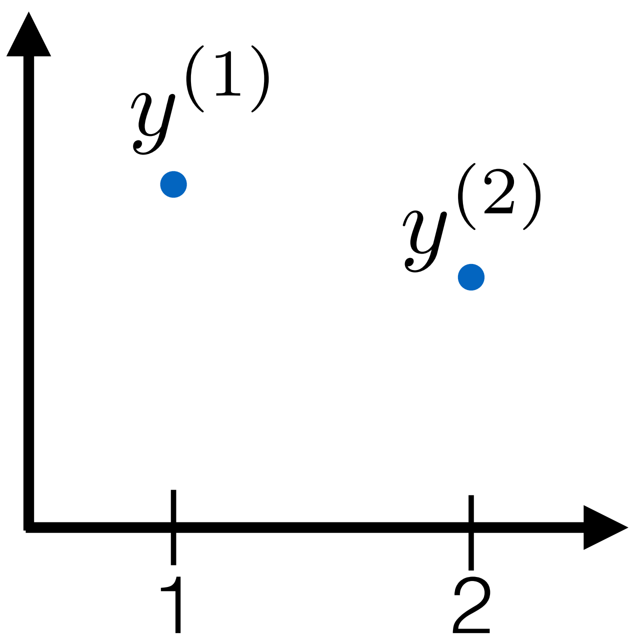


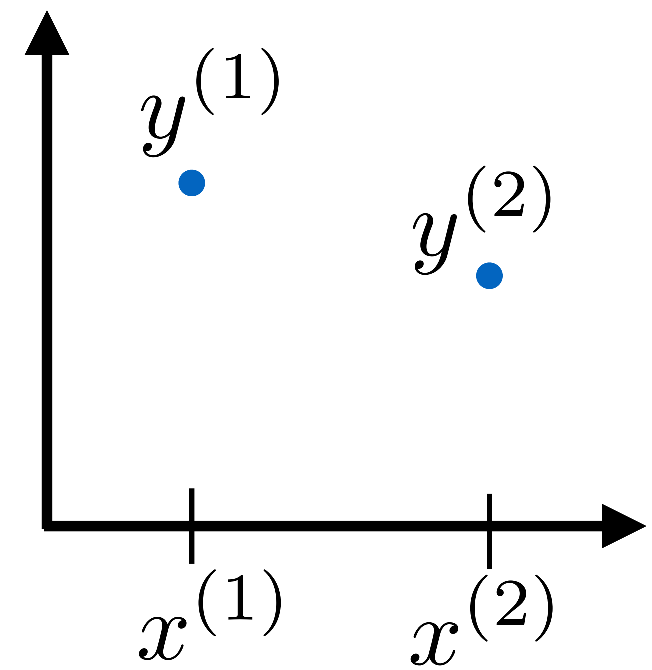


- What if we let the correlation depend on the *x*'s?
  - Let $\rho = \rho(|x^{(1)} - x^{(2)}|)$
    - Where the correlation goes to 1 as the *x*'s get close

Multivariate Gaussian using locations

- *M*=2 dimensions: $y = [y_1, y_2]^\top \sim \mathcal{N}(\mu, K)$
  - With $\mu = [\mu_1, \mu_2]^\top$ and $K = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ ← correlation

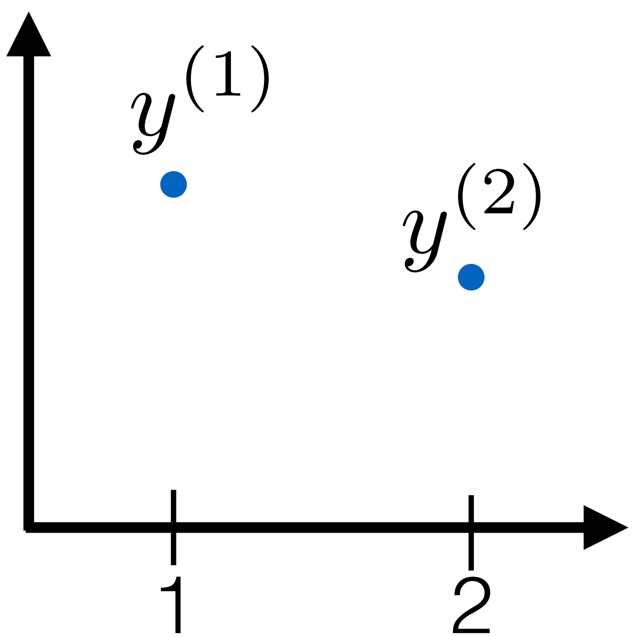


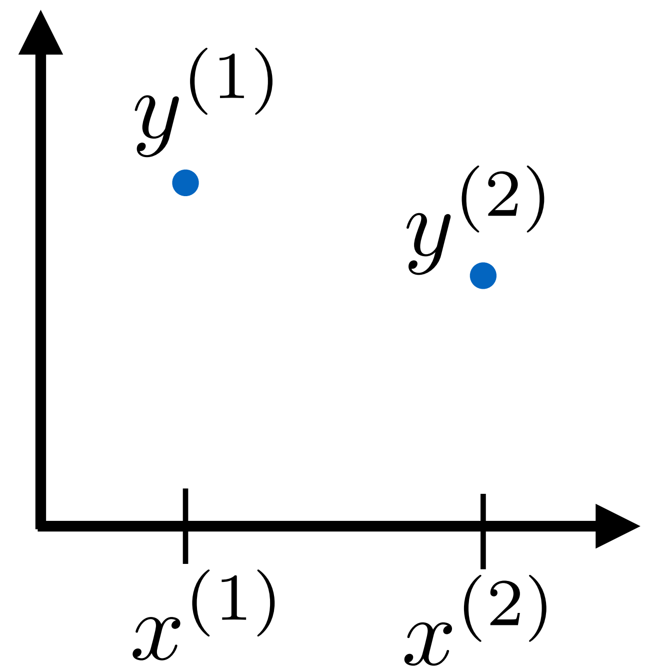


- What if we let the correlation depend on the *x*'s?
  - Let $\rho = \rho(|x^{(1)} - x^{(2)}|)$
    - Where the correlation goes to 1 as the *x*'s get close
    - And goes to 0 as the *x*'s get far

Multivariate Gaussian using locations

- *M*=2 dimensions: $y = [y_1, y_2]^\top \sim \mathcal{N}(\mu, K)$
  - With $\mu = [\mu_1, \mu_2]^\top$ and $K = \sigma^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ ← correlation





- What if we let the correlation depend on the *x*'s?
  - Let $\rho = \rho(|x^{(1)} - x^{(2)}|)$
    - Where the correlation goes to 1 as the *x*'s get close
    - And goes to 0 as the *x*'s get far

[demo]

Multivariate Gaussian using locations

Multivariate Gaussian using locations

- Next: Similar setup but an *M*-dimensional Gaussian instead of just 2 dimensions

Multivariate Gaussian using locations

- Next: Similar setup but an *M*-dimensional Gaussian instead of just 2 dimensions
  - We have *M* locations (need not be in order):

$$x^{(m)} \in (-\infty, \infty)$$

Multivariate Gaussian using locations

- Next: Similar setup but an *M*-dimensional Gaussian instead of just 2 dimensions
  - We have *M* locations (need not be in order):

$$x^{(m)} \in (-\infty, \infty)$$

  - We're going to generate a random $y = [y^{(1)}, \ldots, y^{(M)}]^\top$

Multivariate Gaussian using locations

- Next: Similar setup but an *M*-dimensional Gaussian instead of just 2 dimensions
  - We have *M* locations (need not be in order):

$$x^{(m)} \in (-\infty, \infty)$$

  - We're going to generate a random $y = [y^{(1)}, \ldots, y^{(M)}]^\top$
    - Let $y \sim \mathcal{N}(\mu, K)$

Multivariate Gaussian using locations

- Next: Similar setup but an *M*-dimensional Gaussian instead of just 2 dimensions
  - We have *M* locations (need not be in order):
    $$x^{(m)} \in (-\infty, \infty)$$
  - We're going to generate a random $y = [y^{(1)}, \ldots, y^{(M)}]^\top$
    - Let $y \sim \mathcal{N}(\mu, K)$ with $\mu = \mathbf{0}_M$

Multivariate Gaussian using locations

- Next: Similar setup but an *M*-dimensional Gaussian instead of just 2 dimensions
  - We have *M* locations (need not be in order):
$$x^{(m)} \in (-\infty, \infty)$$
  - We're going to generate a random $y = [y^{(1)}, \ldots, y^{(M)}]^\top$
    - Let $y \sim \mathcal{N}(\mu, K)$ with $\mu = \mathbf{0}_M$ and *K* such that

Multivariate Gaussian using locations

- Next: Similar setup but an *M*-dimensional Gaussian instead of just 2 dimensions
  - We have *M* locations (need not be in order):
    $$x^{(m)} \in (-\infty, \infty)$$
  - We're going to generate a random $y = [y^{(1)}, \ldots, y^{(M)}]^\top$
    - Let $y \sim \mathcal{N}(\mu, K)$ with $\mu = \mathbf{0}_M$ and *K* such that
      $$K_{m,m'} = \sigma^2 \rho(|x^{(m)} - x^{(m')}|)$$

Multivariate Gaussian using locations

- Next: Similar setup but an *M*-dimensional Gaussian instead of just 2 dimensions
  - We have *M* locations (need not be in order):
    $$x^{(m)} \in (-\infty, \infty)$$
  - We're going to generate a random $y = [y^{(1)}, \ldots, y^{(M)}]^\top$
    - Let $y \sim \mathcal{N}(\mu, K)$ with $\mu = \mathbf{0}_M$ and *K* such that
      $$K_{m,m'} = \sigma^2 \rho(|x^{(m)} - x^{(m')}|)$$
    - Let's try $\rho(\Delta) = \exp(-\frac{1}{2}\Delta^2)$

Multivariate Gaussian using locations

- Next: Similar setup but an *M*-dimensional Gaussian instead of just 2 dimensions
  - We have *M* locations (need not be in order):
    $$x^{(m)} \in (-\infty, \infty)$$
  - We're going to generate a random $y = [y^{(1)}, \ldots, y^{(M)}]^\top$
    - Let $y \sim \mathcal{N}(\mu, K)$ with $\mu = \mathbf{0}_M$ and *K* such that
      $$K_{m,m'} = \sigma^2 \rho(|x^{(m)} - x^{(m')}|)$$
    - Let's try $\rho(\Delta) = \exp(-\frac{1}{2}\Delta^2)$
      - Check: $\rho(0) =$?

Multivariate Gaussian using locations

- Next: Similar setup but an *M*-dimensional Gaussian instead of just 2 dimensions
  - We have *M* locations (need not be in order):
    $$x^{(m)} \in (-\infty, \infty)$$
  - We're going to generate a random $y = [y^{(1)}, \ldots, y^{(M)}]^\top$
    - Let $y \sim \mathcal{N}(\mu, K)$ with $\mu = \mathbf{0}_M$ and *K* such that
      $$K_{m,m'} = \sigma^2 \rho(|x^{(m)} - x^{(m')}|)$$
    - Let's try $\rho(\Delta) = \exp(-\frac{1}{2}\Delta^2)$
      - Check: $\rho(0) = 1$

Multivariate Gaussian using locations

- Next: Similar setup but an *M*-dimensional Gaussian instead of just 2 dimensions
  - We have *M* locations (need not be in order):
    $$x^{(m)} \in (-\infty, \infty)$$
  - We're going to generate a random $y = [y^{(1)}, \ldots, y^{(M)}]^\top$
    - Let $y \sim \mathcal{N}(\mu, K)$ with $\mu = \mathbf{0}_M$ and *K* such that
      $$K_{m,m'} = \sigma^2 \rho(|x^{(m)} - x^{(m')}|)$$
    - Let's try $\rho(\Delta) = \exp(-\frac{1}{2}\Delta^2)$
      - Check: $\rho(0) = 1$, $\rho(\Delta)$ is ? as Δ increases

Multivariate Gaussian using locations

- Next: Similar setup but an *M*-dimensional Gaussian instead of just 2 dimensions
  - We have *M* locations (need not be in order):
    $$x^{(m)} \in (-\infty, \infty)$$
  - We're going to generate a random $y = [y^{(1)}, \ldots, y^{(M)}]^\top$
    - Let $y \sim \mathcal{N}(\mu, K)$ with $\mu = \mathbf{0}_M$ and *K* such that
      $$K_{m,m'} = \sigma^2 \rho(|x^{(m)} - x^{(m')}|)$$
    - Let's try $\rho(\Delta) = \exp(-\frac{1}{2}\Delta^2)$
      - Check: $\rho(0) = 1$, $\rho(\Delta)$ is decreasing as Δ increases

Multivariate Gaussian using locations

- Next: Similar setup but an *M*-dimensional Gaussian instead of just 2 dimensions
  - We have *M* locations (need not be in order):
    $$x^{(m)} \in (-\infty, \infty)$$
  - We're going to generate a random $y = [y^{(1)}, \ldots, y^{(M)}]^\top$
    - Let $y \sim \mathcal{N}(\mu, K)$ with $\mu = \mathbf{0}_M$ and *K* such that
      $$K_{m,m'} = \sigma^2 \rho(|x^{(m)} - x^{(m')}|)$$
    - Let's try $\rho(\Delta) = \exp(-\frac{1}{2}\Delta^2)$
      - Check: $\rho(0) = 1$, $\rho(\Delta)$ is decreasing as Δ increases
      - And $\rho(\Delta) \to$? as $\Delta \to \infty$

Multivariate Gaussian using locations

- Next: Similar setup but an *M*-dimensional Gaussian instead of just 2 dimensions
  - We have *M* locations (need not be in order):
    $$x^{(m)} \in (-\infty, \infty)$$
  - We're going to generate a random $y = [y^{(1)}, \ldots, y^{(M)}]^\top$
    - Let $y \sim \mathcal{N}(\mu, K)$ with $\mu = \mathbf{0}_M$ and *K* such that
      $$K_{m,m'} = \sigma^2 \rho(|x^{(m)} - x^{(m')}|)$$
    - Let's try $\rho(\Delta) = \exp(-\frac{1}{2}\Delta^2)$
      - Check: $\rho(0) = 1$, $\rho(\Delta)$ is decreasing as Δ increases
      - And $\rho(\Delta) \rightarrow 0$ as $\Delta \rightarrow \infty$

Multivariate Gaussian using locations

- Next: Similar setup but an *M*-dimensional Gaussian instead of just 2 dimensions
  - We have *M* locations (need not be in order):
    $$x^{(m)} \in (-\infty, \infty)$$
  - We're going to generate a random $y = [y^{(1)}, \ldots, y^{(M)}]^\top$
    - Let $y \sim \mathcal{N}(\mu, K)$ with $\mu = \mathbf{0}_M$ and *K* such that
      $$K_{m,m'} = \sigma^2 \rho(|x^{(m)} - x^{(m')}|)$$
    - Let's try $\rho(\Delta) = \exp(-\frac{1}{2}\Delta^2)$
      - Check: $\rho(0) = 1$, $\rho(\Delta)$ is decreasing as Δ increases
      - And $\rho(\Delta) \to 0$ as $\Delta \to \infty$

Note: our previous example was the special case where *M*=2

Multivariate Gaussian using locations

- Next: Similar setup but an *M*-dimensional Gaussian instead of just 2 dimensions
  - We have *M* locations (need not be in order):
    $$x^{(m)} \in (-\infty, \infty)$$
  - We're going to generate a random $y = [y^{(1)}, \ldots, y^{(M)}]^\top$
    - Let $y \sim \mathcal{N}(\mu, K)$ with $\mu = \mathbf{0}_M$ and *K* such that
      $$K_{m,m'} = \sigma^2 \rho(|x^{(m)} - x^{(m')}|)$$
    - Let's try $\rho(\Delta) = \exp(-\frac{1}{2}\Delta^2)$
      - Check: $\rho(0) = 1$, $\rho(\Delta)$ is decreasing as Δ increases
      - And $\rho(\Delta) \to 0$ as $\Delta \to \infty$

[demo1, demo2]

Note: our previous example was the special case where *M*=2

Multivariate Gaussian using locations

- Next: Similar setup but an *M*-dimensional Gaussian instead of just 2 dimensions
  - We have *M* locations (need not be in order):
$$x^{(m)} \in (-\infty, \infty)$$
  - We're going to generate a random $y = [y^{(1)}, \ldots, y^{(M)}]^\top$
    - Let $y \sim \mathcal{N}(\mu, K)$ with $\mu = \mathbf{0}_M$ and *K* such that
$$K_{m,m'} = \sigma^2 \rho(|x^{(m)} - x^{(m')}|)$$
    - Let's try $\rho(\Delta) = \exp(-\frac{1}{2}\Delta^2)$
      - Check: $\rho(0) = 1$, $\rho(\Delta)$ is decreasing as Δ increases
      - And $\rho(\Delta) \to 0$ as $\Delta \to \infty$

[demo1, demo2]

Note: our previous example was the special case where *M*=2

We just drew random functions from a type of Gaussian process that is very commonly used in practice!

Gaussian processes

Gaussian processes

- Definition: “A *Gaussian process* is a collection of random variables, any finite number of which have a joint Gaussian distribution.” [Rasmussen and Williams 2006; a much much older idea!]

Gaussian processes

- Definition: "A *Gaussian process* is a collection of random variables, any finite number of which have a joint Gaussian distribution." [Rasmussen and Williams 2006; a much much older idea!]
  - E.g. the function $f(\mathbf{x})$ is a collection indexed by input $\mathbf{x}$

Gaussian processes

- Definition: “A *Gaussian process* is a collection of random variables, any finite number of which have a joint Gaussian distribution.” [Rasmussen and Williams 2006; a much much older idea!]
  - E.g. the function $f(\mathbf{x})$ is a collection indexed by input $\mathbf{x}$
- It is specified by its mean function and covariance function:

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