Gaussian Processes

Mark van der Wilk

Department of Computing
Imperial College London

@markvanderwilk
m.vdwilk@imperial.ac.uk

January 27, 2023

Recap: Model

- Specify prior on weights $p(\mathbf{w})$

Recap: Model

- Specify prior on weights $p(\mathbf{w})$
- Defines distribution on functions through $f(\mathbf{x}) = \boldsymbol{\phi}(\mathbf{x})^\top \mathbf{w}$

Recap: Model

- Specify prior on weights $p(\mathbf{w})$
- Defines distribution on functions through $f(\mathbf{x}) = \boldsymbol{\phi}(\mathbf{x})^\intercal \mathbf{w}$
- Observe data through likelihood $p(\mathbf{y}|f(X)) = \prod_{n=1}^{N} \mathcal{N}\left(y_n; f(\mathbf{x}_n), \sigma^2\right)$

Recap: Inference

- Find posterior on weights $p(\mathbf{w}|\mathbf{y})$

Recap: Inference

- Find posterior on weights $p(\mathbf{w}|\mathbf{y})$
- Combine this with $p(\mathbf{y}^*|\mathbf{w})$ to find $p(\mathbf{y}^*|\mathbf{y})$:

$$p(\mathbf{y}_*|\mathbf{y}) = \mathcal{N}\Big(\boldsymbol{\theta};\quad \boldsymbol{\phi}(\mathbf{x}_*)^\top\big[\boldsymbol{I}_M + \sigma^{-2}\Phi(\boldsymbol{X})^\top\Phi(\boldsymbol{X})\big]^{-1}\sigma^{-2}\Phi(\boldsymbol{X})^\top\mathbf{y}$$
$$\boldsymbol{\phi}(\mathbf{x}_*)^\top\big[\boldsymbol{I}_M + \sigma^{-2}\Phi(\boldsymbol{X})^\top\Phi(\boldsymbol{X})\big]^{-1}\boldsymbol{\phi}(\mathbf{x}_*) + \sigma^2\boldsymbol{I}_N\Big)$$

Recap: Inference

- Find posterior on weights $p(\mathbf{w}|\mathbf{y})$
- Combine this with $p(\mathbf{y}^*|\mathbf{w})$ to find $p(\mathbf{y}^*|\mathbf{y})$:

$$p(\mathbf{y}_*|\mathbf{y}) = \mathcal{N}\Big(\boldsymbol{\theta}; \quad \boldsymbol{\phi}(\mathbf{x}_*)^\top \big[\boldsymbol{I}_M + \sigma^{-2}\Phi(\boldsymbol{X})^\top\Phi(\boldsymbol{X})\big]^{-1}\sigma^{-2}\Phi(\boldsymbol{X})^\top\mathbf{y}$$
$$\boldsymbol{\phi}(\mathbf{x}_*)^\top \big[\boldsymbol{I}_M + \sigma^{-2}\Phi(\boldsymbol{X})^\top\Phi(\boldsymbol{X})\big]^{-1}\boldsymbol{\phi}(\mathbf{x}_*) + \sigma^2\boldsymbol{I}_N\Big)$$

- Apply Woodbury to go from $O(NM^2 + M^3) \to O(N^3)$:

$$p(y_*|\mathbf{y}) = \mathcal{N}\Big(y_*; \quad \boldsymbol{\phi}(\mathbf{x}_*)^\top\Phi(\boldsymbol{X})^\top\big[\Phi(\boldsymbol{X})\Phi(\boldsymbol{X})^\top + \sigma^2\boldsymbol{I}_N\big]^{-1}\mathbf{y}\,,$$
$$\boldsymbol{\phi}(\mathbf{x}_*)^\top\boldsymbol{\phi}(\mathbf{x}_*) + \sigma^2$$
$$- \boldsymbol{\phi}(\mathbf{x}_*)^\top\Phi(\boldsymbol{X})^\top\big[\Phi(\boldsymbol{X})\Phi(\boldsymbol{X})^\top + \sigma^2\boldsymbol{I}_N\big]^{-1}\Phi(\boldsymbol{X})\boldsymbol{\phi}(\mathbf{x}_*)\Big)$$

Recap: Inference

- Find posterior on weights $p(\mathbf{w}|\mathbf{y})$
- Combine this with $p(\mathbf{y}^*|\mathbf{w})$ to find $p(\mathbf{y}^*|\mathbf{y})$:

$$p(\mathbf{y}_*|\mathbf{y}) = \mathcal{N}\Big(\boldsymbol{\theta};\quad \boldsymbol{\phi}(\mathbf{x}_*)^\top\big[\boldsymbol{I}_M + \sigma^{-2}\Phi(\boldsymbol{X})^\top\Phi(\boldsymbol{X})\big]^{-1}\sigma^{-2}\Phi(\boldsymbol{X})^\top\mathbf{y}$$
$$\boldsymbol{\phi}(\mathbf{x}_*)^\top\big[\boldsymbol{I}_M + \sigma^{-2}\Phi(\boldsymbol{X})^\top\Phi(\boldsymbol{X})\big]^{-1}\boldsymbol{\phi}(\mathbf{x}_*) + \sigma^2\boldsymbol{I}_N\Big)$$

- Apply Woodbury to go from $O(NM^2 + M^3) \to O(N^3)$:
- Apply Kernel trick $\boldsymbol{\phi}(\mathbf{x})^\top\boldsymbol{\phi}(\mathbf{x}') = k(\mathbf{x}, \mathbf{x}')$

$$p(y_*|\mathbf{y}) = \mathcal{N}\Big(y_*;\quad k(\mathbf{x}_*, \boldsymbol{X})\big[k(\boldsymbol{X}, \boldsymbol{X}) + \sigma^2\boldsymbol{I}_N\big]^{-1}\mathbf{y}\,,$$
$$k(\mathbf{x}_*, \mathbf{x}_*) + \sigma^2 - k(\mathbf{x}_*, \boldsymbol{X})\big[k(\boldsymbol{X}, \boldsymbol{X}) + \sigma^2\boldsymbol{I}_N\big]^{-1}k(\boldsymbol{X}, \mathbf{x}_*)\Big)$$

Today

- Develop interpretation of the maths we got from Woodbury

Today

- Develop interpretation of the maths we got from Woodbury
- This is a way of specifying distributions on functions

Today

- Develop interpretation of the maths we got from Woodbury
- This is a way of specifying distributions on functions
- But without parameters!

How do we get rid of parameters?

Model:

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}; 0, \boldsymbol{I}_M) \tag{1}$$

$$f(\mathbf{x}) = \boldsymbol{\phi}(\mathbf{x})^\mathsf{T}\mathbf{w} \tag{2}$$

$$p(\mathbf{y}|f(X)) = \prod_{n=1}^{N} \mathcal{N}\left(y_n; f(\mathbf{x}_n), \sigma^2\right) \tag{3}$$

How do we get rid of parameters?

Model:

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}; 0, \boldsymbol{I}_M) \tag{1}$$

$$f(\mathbf{x}) = \boldsymbol{\phi}(\mathbf{x})^\top \mathbf{w} \tag{2}$$

$$p(\mathbf{y}|f(X)) = \prod_{n=1}^{N} \mathcal{N}\left(y_n; f(\mathbf{x}_n), \sigma^2\right) \tag{3}$$

Observation: Likelihood only depends on function value

How do we get rid of parameters?

Model:

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}; 0, \boldsymbol{I}_M) \tag{1}$$

$$f(\mathbf{x}) = \boldsymbol{\phi}(\mathbf{x})^\top \mathbf{w} \tag{2}$$

$$p(\mathbf{y}|f(X)) = \prod_{n=1}^{N} \mathcal{N}\left(y_n; f(\mathbf{x}_n), \sigma^2\right) \tag{3}$$

Observation: Likelihood only depends on function value
$\implies$ Can we ignore the distribution on weights, and work directly with function values?

How do we get rid of parameters?

Model:

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}; 0, \boldsymbol{I}_M) \tag{1}$$

$$f(\mathbf{x}) = \boldsymbol{\phi}(\mathbf{x})^\mathsf{T}\mathbf{w} \tag{2}$$

$$p(\mathbf{y}|f(X)) = \prod_{n=1}^{N} \mathcal{N}\left(y_n; f(\mathbf{x}_n), \sigma^2\right) \tag{3}$$

Observation: Likelihood only depends on function value
$\implies$ Can we ignore the distribution on weights, and work directly with function values?

All we are **really** interested in, is:

How do we get rid of parameters?

Model:

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}; 0, \boldsymbol{I}_M) \tag{1}$$

$$f(\mathbf{x}) = \boldsymbol{\phi}(\mathbf{x})^\top \mathbf{w} \tag{2}$$

$$p(\mathbf{y}|f(X)) = \prod_{n=1}^{N} \mathcal{N}\left(y_n; f(\mathbf{x}_n), \sigma^2\right) \tag{3}$$

Observation: Likelihood only depends on function value
$\implies$ Can we ignore the distribution on weights, and work directly with function values?

All we are **really** interested in, is:

- Predicting data $p(\mathbf{y}^*|\mathbf{y})$

How do we get rid of parameters?

Model:

$$p(\mathbf{w}) = \mathcal{N}(\mathbf{w}; 0, \boldsymbol{I}_M) \tag{1}$$

$$f(\mathbf{x}) = \boldsymbol{\phi}(\mathbf{x})^\mathsf{T}\mathbf{w} \tag{2}$$

$$p(\mathbf{y}|f(X)) = \prod_{n=1}^{N} \mathcal{N}\left(y_n; f(\mathbf{x}_n), \sigma^2\right) \tag{3}$$

Observation: Likelihood only depends on function value
$\implies$ Can we ignore the distribution on weights, and work directly with function values?

All we are **really** interested in, is:

- Predicting data $p(\mathbf{y}^*|\mathbf{y})$
- Predicting function values $p(f(X^*))|\mathbf{y})$

Distribution on Function Values

Let's start by analysing distribution on function values $p(f(X))$ (board)

Distribution on Function Values

Let's start by analysing distribution on function values $p(f(X))$ (board)

- Function values are linear transformation of Gaussian RV
  For arbitrary N inputs arranged in a matrix $X \in \mathbb{R}^{N \times D}$:

$$f(X) = \Phi(X)\mathbf{w} \tag{4}$$

Distribution on Function Values

Let's start by analysing distribution on function values $p(f(X))$ (board)

- Function values are linear transformation of Gaussian RV
  For arbitrary N inputs arranged in a matrix $X \in \mathbb{R}^{N \times D}$:

$$f(X) = \Phi(X)\mathbf{w} \tag{4}$$

- As usual $\implies$ Gaussian distributed, and can find mean+var

Distribution on Function Values

Let's start by analysing distribution on function values $p(f(X))$ (board)

- Function values are linear transformation of Gaussian RV
  For arbitrary N inputs arranged in a matrix $X \in \mathbb{R}^{N \times D}$:

$$f(X) = \Phi(X)\mathbf{w} \tag{4}$$

- As usual $\implies$ Gaussian distributed, and can find mean+var

$$\mathbb{E}_{\mathbf{w}}[f(X)] = \mathbb{E}_{\mathbf{w}}[\Phi(X)\mathbf{w}] = 0 \tag{5}$$

$$\mathbb{V}_{\mathbf{w}}[f(X)] = \mathbb{E}_{\mathbf{w}}[\Phi(X)\mathbf{w}\mathbf{w}^\intercal\Phi(X)^\intercal] = \Phi(X)\Phi(X)^\intercal \tag{6}$$

$$\implies p(f(X)) = \mathcal{N}(f(X); 0, \Phi(X)\Phi(X)^\intercal) \tag{7}$$

Distribution on Function Values

Let's start by analysing distribution on function values $p(f(X))$ (board)

- Function values are linear transformation of Gaussian RV
  For arbitrary N inputs arranged in a matrix $X \in \mathbb{R}^{N \times D}$:

$$f(X) = \Phi(X)\mathbf{w} \tag{4}$$

- As usual $\implies$ Gaussian distributed, and can find mean+var

$$\mathbb{E}_{\mathbf{w}}[f(X)] = \mathbb{E}_{\mathbf{w}}[\Phi(X)\mathbf{w}] = 0 \tag{5}$$

$$\mathbb{V}_{\mathbf{w}}[f(X)] = \mathbb{E}_{\mathbf{w}}[\Phi(X)\mathbf{w}\mathbf{w}^{\mathsf{T}}\Phi(X)^{\mathsf{T}}] = \Phi(X)\Phi(X)^{\mathsf{T}} \tag{6}$$

$$\implies p(f(X)) = \mathcal{N}(f(X); 0, \Phi(X)\Phi(X)^{\mathsf{T}}) \tag{7}$$

- All function values are correlated

Distribution on Function Values

Let's start by analysing distribution on function values $p(f(X))$ (board)

- Function values are linear transformation of Gaussian RV
  For arbitrary N inputs arranged in a matrix $X \in \mathbb{R}^{N \times D}$:

$$f(X) = \Phi(X)\mathbf{w} \tag{4}$$

- As usual $\implies$ Gaussian distributed, and can find mean+var

$$\mathbb{E}_{\mathbf{w}}[f(X)] = \mathbb{E}_{\mathbf{w}}[\Phi(X)\mathbf{w}] = 0 \tag{5}$$

$$\mathbb{V}_{\mathbf{w}}[f(X)] = \mathbb{E}_{\mathbf{w}}[\Phi(X)\mathbf{w}\mathbf{w}^\top\Phi(X)^\top] = \Phi(X)\Phi(X)^\top \tag{6}$$

$$\implies p(f(X)) = \mathcal{N}(f(X); 0, \Phi(X)\Phi(X)^\top) \tag{7}$$

- All function values are correlated
- Kernel trick applies! $[\Phi(X)\Phi(X)^\top]_{ij} = \boldsymbol{\phi}(\mathbf{x}_i)^\top\boldsymbol{\phi}(\mathbf{x}_j) = k(\mathbf{x}_i, \mathbf{x}_j)$

Predicting

Let's focus on $p(f(X^*)|\mathbf{y})$.

$$p(f(X^*)|\mathbf{y}) \overset{\text{AT}}{=} \frac{\int p(\mathbf{y}, f(X), f(X^*))\ \mathrm{d}f(X)}{p(\mathbf{y})}$$
$$\underset{=}{\text{MA}} \frac{\int p(\mathbf{y}|f(X))\, p(f(X), f(X^*))\ \mathrm{d}f(X)}{p(\mathbf{y})}$$

Predicting

Let's focus on $p(f(X^*)|\mathbf{y})$.

$$p(f(X^*)|\mathbf{y}) \overset{\text{AT}}{=} \frac{\int p(\mathbf{y}, f(X), f(X^*))\, \mathrm{d}f(X)}{p(\mathbf{y})}$$

$$\overset{\text{MA}}{=} \frac{\int p(\mathbf{y}|f(X))\, p(f(X), f(X^*))\, \mathrm{d}f(X)}{p(\mathbf{y})}$$

$$p(f(X), f(X^*)) = \mathcal{N}\left(\begin{bmatrix} f(X) \\ f(X^*) \end{bmatrix}; 0, \begin{bmatrix} \Phi(X)\Phi(X)^\mathsf{T} & \Phi(X)\Phi(X^*) \\ \Phi(X^*)\Phi(X)^\mathsf{T} & \Phi(X^*)\Phi(X^*)^\mathsf{T} \end{bmatrix}\right)$$

Predicting

Let's focus on $p(f(X^*)|\mathbf{y})$.

$$p(f(X^*)|\mathbf{y}) \overset{\text{AT}}{=} \frac{\int p(\mathbf{y}, f(X), f(X^*)) \, \mathrm{d}f(X)}{p(\mathbf{y})}$$
$$\underset{=}{\text{MA}} \frac{\int p(\mathbf{y}|f(X)) \, p(f(X), f(X^*)) \, \mathrm{d}f(X)}{p(\mathbf{y})}$$

$$p(f(X), f(X^*)) = \mathcal{N}\left(\begin{bmatrix} f(X) \\ f(X^*) \end{bmatrix}; 0, \begin{bmatrix} k(X,X) & k(X,X^*) \\ k(X^*,X) & k(X^*,X^*) \end{bmatrix}\right)$$

Predicting

Let's focus on $p(f(X^*)|\mathbf{y})$.

$$p(f(X^*)|\mathbf{y}) \overset{\text{AT}}{=} \frac{\int p(\mathbf{y}, f(X), f(X^*))\, \mathrm{d}f(X)}{p(\mathbf{y})}$$

$$\overset{\text{MA}}{=} \frac{\int p(\mathbf{y}|f(X))\, p(f(X), f(X^*))\, \mathrm{d}f(X)}{p(\mathbf{y})}$$

$$p(f(X), f(X^*)) = \mathcal{N}\left(\begin{bmatrix} f(X) \\ f(X^*) \end{bmatrix}; 0, \begin{bmatrix} k(X, X) & k(X, X^*) \\ k(X^*, X) & k(X^*, X^*) \end{bmatrix}\right)$$

Easiest way: Find joint, Gaussian conditioning (board)

$$\mathcal{N}\left(\begin{bmatrix} f(X^*) \\ \mathbf{y} \end{bmatrix}; 0, \begin{bmatrix} k(X^*, X^*) & k(X^*, X) \\ k(X, X^*) & k(X, X) + \sigma^2 \boldsymbol{I}_N \end{bmatrix}\right) \tag{8}$$

Predicting

Let's focus on $p(f(X^*)|\mathbf{y})$.

$$p(f(X^*)|\mathbf{y}) \overset{\text{AT}}{=} \frac{\int p(\mathbf{y}, f(X), f(X^*)) \, \mathrm{d}f(X)}{p(\mathbf{y})}$$

$$\overset{\text{MA}}{=} \frac{\int p(\mathbf{y}|f(X)) \, p(f(X), f(X^*)) \, \mathrm{d}f(X)}{p(\mathbf{y})}$$

$$p(f(X), f(X^*)) = \mathcal{N}\left(\begin{bmatrix} f(X) \\ f(X^*) \end{bmatrix}; 0, \begin{bmatrix} k(X, X) & k(X, X^*) \\ k(X^*, X) & k(X^*, X^*) \end{bmatrix}\right)$$

Easiest way: Find joint, Gaussian conditioning (board)

$$\mathcal{N}\left(\begin{bmatrix} f(X^*) \\ \mathbf{y} \end{bmatrix}; 0, \begin{bmatrix} k(X^*, X^*) & k(X^*, X) \\ k(X, X^*) & k(X, X) + \sigma^2 \boldsymbol{I}_N \end{bmatrix}\right) \tag{8}$$

$$p(f(X^*)|\mathbf{y}) = \mathcal{N}\Big(f(X^*); \quad k(X^*, X)\left[k(X, X) + \sigma^2 \boldsymbol{I}_N\right]^{-1} \mathbf{y},$$
$$k(X^*, X^*) - k(X^*, X)\left[k(X, X) + \sigma^2 \boldsymbol{I}_N\right]^{-1} k(X, X^*)\Big)$$

What is a Gaussian Process?

- Same as what we get from BLR + Woodbury + kernel trick!

What is a Gaussian Process?

- Same as what we get from BLR + Woodbury + kernel trick!
- No need for parameters, only kernel k!

What is a Gaussian Process?

- Same as what we get from BLR + Woodbury + kernel trick!
- No need for parameters, only kernel k!

Who needs parameters?

What is a Gaussian Process?

- Same as what we get from BLR + Woodbury + kernel trick!
- No need for parameters, only kernel k!

Who needs parameters?
$\implies$ Can answer any prediction question
using only distribution on function *values*.

What is a Gaussian Process?

A (possibly infinite) collection of Random Variables

What is a Gaussian Process?

A (possibly infinite) collection of Random Variables
such that each finite collection has a Gaussian distribution.

What is a Gaussian Process?

A (possibly infinite) collection of Random Variables
such that each finite collection has a Gaussian distribution.

Properties

- I will index this collection with x.
- For this to be a valid collection of RVs, sum rule must hold:

What is a Gaussian Process?

A (possibly infinite) collection of Random Variables
such that each finite collection has a Gaussian distribution.

Properties

- I will index this collection with x.
- For this to be a valid collection of RVs, sum rule must hold:

$$p(f(x_1), f(x_2)) = \int p(f(x_1), f(x_2), f(x_3)) \mathrm{d}f(x_3) \tag{9}$$

Specifying Gaussian Processes

Can specify the function value densities $p(f(X))$ using:

- Mean function $\mu : \mathcal{X} \to \mathbb{R}$
- Covariance function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

Specifying Gaussian Processes

Can specify the function value densities $p(f(X))$ using:

- Mean function $\mu : \mathcal{X} \to \mathbb{R}$
- Covariance function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

$$p(f(X)) = \mathcal{N}(f(X); \mu(X), k(X, X))$$

$$\mu(X) \in \mathbb{R}^N \qquad k(X, X) \in \mathbb{R}^{D \times D}$$

$$[\mu(X)]_i = \mu(\mathbf{x}_i) \qquad [k(X, X)]_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$$

Specifying Gaussian Processes

Can specify the function value densities $p(f(X))$ using:

- Mean function $\mu : \mathcal{X} \to \mathbb{R}$
- Covariance function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$

$$p(f(X)) = \mathcal{N}(f(X); \mu(X), k(X, X))$$

$$\mu(X) \in \mathbb{R}^N \qquad\qquad k(X, X) \in \mathbb{R}^{D\times D}$$

$$[\mu(X)]_i = \mu(\mathbf{x}_i) \qquad\qquad [k(X, X)]_{ij} = k(\mathbf{x}_i, \mathbf{x}_j)$$

Covariance function $k(\cdot, \cdot)$ must be a positive definite function.
I.e. $k(X, X)$ is PSD for any choice of X.

BLR Specifies a GP

- BLR specifies a density of a collection of RVs

BLR Specifies a GP

- BLR specifies a density of a collection of RVs
- Collection of random variables is **function values** at all locations

$$p(f(X)) = \mathcal{N}(f(X); 0, \Phi(X)\Phi(X)^\top) \tag{10}$$

$$= \mathcal{N}(f(X); 0, k(X, X)) \tag{11}$$

BLR Specifies a GP

- BLR specifies a density of a collection of RVs
- Collection of random variables is **function values** at all locations

$$
\begin{aligned}
p(f(X)) &= \mathcal{N}(f(X); 0, \Phi(X)\Phi(X)^\top) && (10)\\
&= \mathcal{N}(f(X); 0, k(X, X)) && (11)
\end{aligned}
$$

- $\implies$ BLR specifies a GP and a kernel

BLR Specifies a GP

- BLR specifies a density of a collection of RVs
- Collection of random variables is **function values** at all locations

$$p(f(X)) = \mathcal{N}(f(X); 0, \Phi(X)\Phi(X)^\top) \tag{10}$$
$$= \mathcal{N}(f(X); 0, k(X, X)) \tag{11}$$

- $\implies$ BLR specifies a GP and a kernel
- Directly specifying a kernel, also specifies a GP

BLR Specifies a GP

- BLR specifies a density of a collection of RVs
- Collection of random variables is **function values** at all locations

$$p(f(X)) = \mathcal{N}(f(X); 0, \Phi(X)\Phi(X)^\top) \tag{10}$$
$$= \mathcal{N}(f(X); 0, k(X, X)) \tag{11}$$

- $\implies$ BLR specifies a GP and a kernel
- Directly specifying a kernel, also specifies a GP
- We viewed BLR as specifying a distribution on functions

GPs as distributions on functions

Can we view a GP as a distribution on functions?

GPs as distributions on functions

Can we view a GP as a distribution on functions?
▶▶ Yes! Kolmogorov Extension Theorem (not examined).

GPs as distributions on functions

Can we view a GP as a distribution on functions?
▶▶ Yes! Kolmogorov Extension Theorem (not examined).

Conclusion

- Covariance functions / kernels specify GPs

Conclusion

- Covariance functions / kernels specify GPs
- GPs specify distributions on function values directly

Conclusion

- Covariance functions / kernels specify GPs
- GPs specify distributions on function values directly
- To make predictions, we only need distributions on function values

Conclusion

- Covariance functions / kernels specify GPs
- GPs specify distributions on function values directly
- To make predictions, we only need distributions on function values
- So who needs parameters?

Conclusion

- Covariance functions / kernels specify GPs
- GPs specify distributions on function values directly
- To make predictions, we only need distributions on function values
- So who needs parameters?
- BLR specifies a GP

Recommended reading

- Rasmussen and Williams (2006) §2.1 + §2.2

References I

Rasmussen, C. E. and Williams, C. K. (2006). *Gaussian processes for machine learning*. MIT press, Cambridge, MA, USA.