Machine Learning for Computer Graphics

Previous Classes

Basics of Ray Tracing and Phong Illumination Model

Basic Radiometric Quantities, BRDF and Illumination Integral

Monte Carlo Integration

Monte Carlo Integration for Illumination Integrals (i.e., for rendering)

This Class

Identify the sources of estimation error in Monte Carlo Integration

Question the fundamentals of Monte Carlo Integration

Bayesian Monte Carlo: a Machine Learning-based alternative

- Leverages Gaussian Process Regression (GPR)

Course Roadmap

Week no.	Class Date	Theory	Labs	Notes
Week 1	22/02/2022	1.1 Intro	-	-
Week 2	02/03/2022	1.2 Ray-Tracing Phong	Lab 1	P1 Phong
Week 3	08/03/2022	1.3 Radiometry and BRDF	Lab 2	
Week 4	15/03/2022	2.1 Monte Carlo (MC)	Lab 3	P2 MC
Week 5	22/03/2022	2.2 MC Rendering	Lab 4	
Week 6	**29/03/2022**	**2.3 Bayesian Monte Carlo (BMC)**	**Lab 5**	**P3 BMC**
	05/04/2022	PARTIAL EXAMS (NO CLASS)		
	12/04/2022	EASTER HOLYDAYS (NO CLASS)		
Week 7	19/04/2022	2.4 BMC Rendering	Lab 6	
Week 8	26/04/2022	3.1 MC Importance Sampling	Lab 7	P4 AMC
Week 9	03/05/2022	3.2 BMC Importance Sampling	Lab 8	Papers Out
	10/05/2022	AI SEMINAR MAI (NO CLASS)		
Week 10	17/05/2022	Presentations I		
Week 11	24/05/2022	Presentations II		

Preamble

About Monte Carlo Integration

Challenging the MC approach

$$\hat{I} = \frac{1}{N}\sum_{j=1}^{N}\frac{f(x_j)}{p(x_j)}$$

"Monte Carlo is Fundamentally Unsound" - (O'Hagan, 1987)

O'Hagan presents two main objections to the MC method

- Objection 1: The estimate depends on chosen pdf
- Objection 2: Monte Carlo ignores inter-samples distance

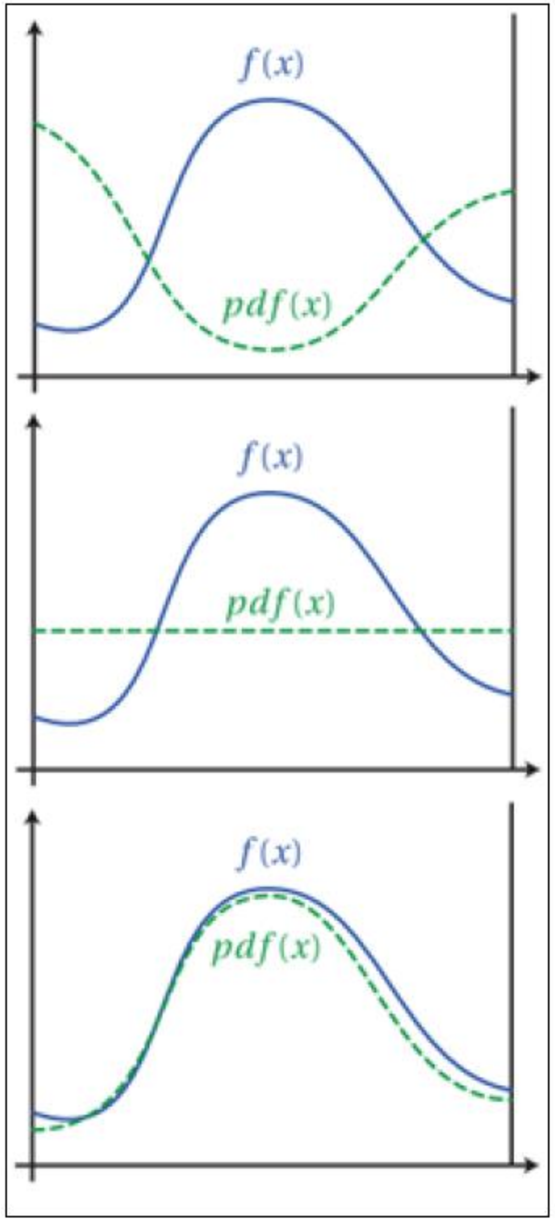
Objection 1

"The estimate depends on the chosen pdf"

$$\hat{I} = \frac{1}{N}\sum_{j=1}^{N}\frac{f(x_j)}{p(x_j)}$$

The pdf choice can be arbitrary, or based on the practitioner's beliefs

How does this choice affect the value of a particular estimate?



Objection 1

"The estimate depends on the chosen pdf"

Practical example:

- Consider two *different* pdfs: $\mathrm{pdf}_1(x)$ and $\mathrm{pdf}_2(x)$
- We take 3 samples from each pdf
- And, by luck, we get the *same* samples $\{f(x_1), f(x_2), f(x_3)\}$

Since $\mathrm{pdf}_1(x) \neq \mathrm{pdf}_2(x)$, we have two different estimates

$$\hat{I} = \frac{1}{N}\sum_{j=1}^{N}\frac{f(x_j)}{\mathrm{pdf}(x_j)}$$

Same observations, two different results... Reasonable?

Objection 2

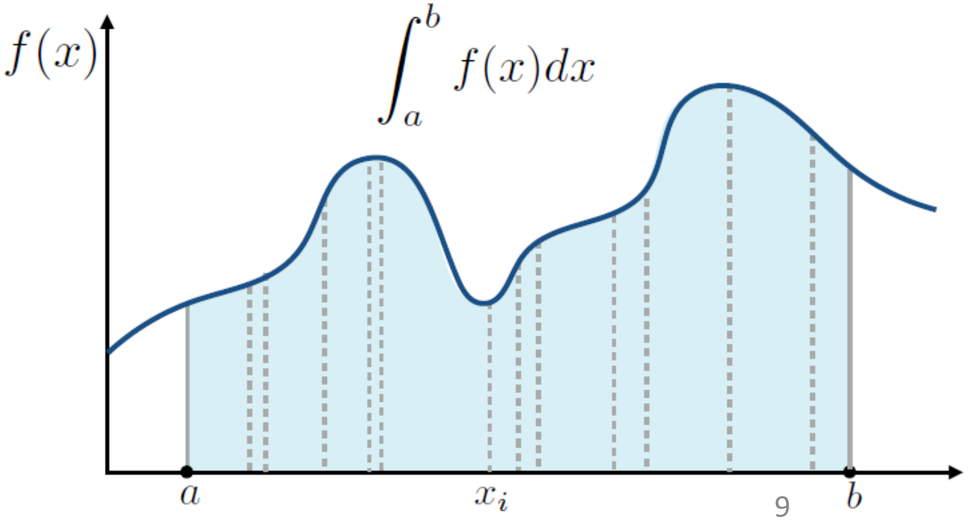
$$\hat{I} = \frac{1}{N}\sum_{j=1}^{N}\frac{f(x_j)}{p(x_j)}$$

"Monte Carlo ignores inter samples distance"

The Monte Carlo method does not account for the distance between the samples

Nearby samples might carry redundant information

An isolated sample might be the single source of information in a particular region of our integrand



Objection 2 (Weights View)

$$\hat{I} = \frac{1}{N}\sum_{j=1}^{N}\frac{f(x_j)}{p(x_j)}$$

Let us consider the MC estimator $\hat{I}$ of the integral $I = \int_a^b f(x)\,\mathrm{d}x$

Let $p(x)$ be a uniform pdf given by $p(x) = \frac{1}{b-a}$

We can write $\hat{I} = \sum_{n=1}^{N} w_n\, f(x_n)$

- $w_n = \frac{b-a}{N}$ is the weight of each sample
- The weight of a particular sample is independent of the distance to other samples (practical example in next slide)

Objection 2

$$\hat{I} = \frac{1}{N}\sum_{j=1}^{N}\frac{f(x_j)}{p(x_j)}$$

Monte Carlo ignores inter samples distance

Practical example:

- Consider three samples $\{f(\omega_1), f(\omega_2), f(\omega_3)\}$
- Suppose $\omega_3 = \omega_2$
- MC estimate using $\{f(\omega_1), f(\omega_2)\}$ and $\{f(\omega_1), f(\omega_2), f(\omega_3)\}$

No new information results in a different result... Reasonable?

Partial Conclusions

Objection 1: The estimate depends on chosen pdf

Objection 2: Monte Carlo ignores inter samples distance

MC ignores relevant information

- Negative impact in the quality of the final estimates
- Even more critical in rendering (samples are scarce)

Goal: method which makes the most of the information available

Bayesian Monte Carlo

Provides a sound framework for stating prior knowledge

Uses all information available

- For example, inter-samples distance

Resorts to Gaussian Process **regression**

- Use the data (i.e., the samples) to learn a probabilistic model representation of the unknown function

Introduction to Linear Models for Regression

Based on Chapters 1 and 3 of Bishop 2006

Linear Regression (LR)

- Problem formulation

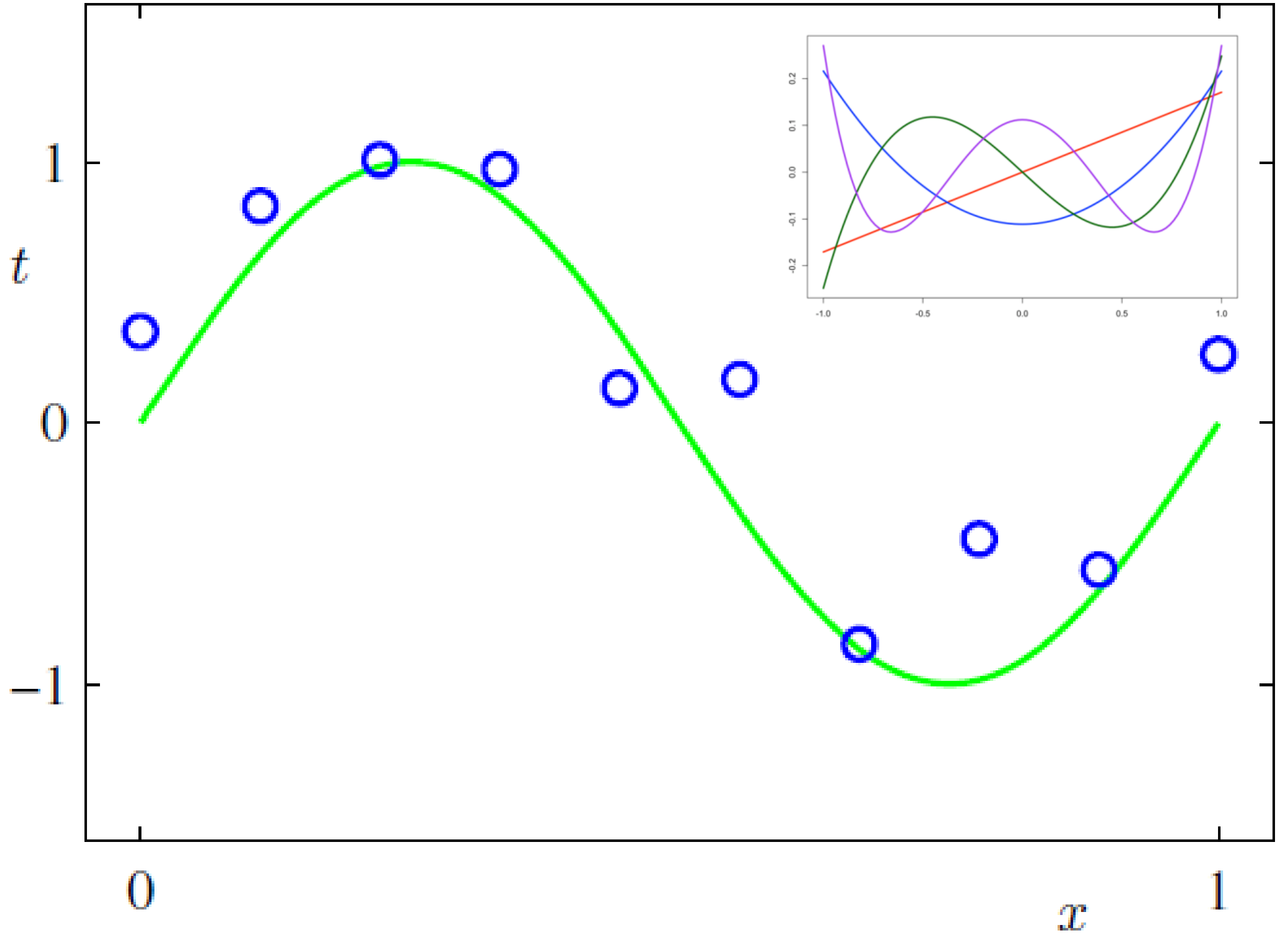
$$D = \{(x_i, y_i) \mid i = 1, \dots, n\}$$

$$y_i = f(x_i) \approx m(x_i) = \boldsymbol{w}^t \boldsymbol{\Phi}(x_i)$$

- Example: polynomial basis

$$\Phi(x) = [\Phi_1(x), \dots, \Phi_M(x)]$$
$$= [1, x, \dots, x^{M-1}]$$

- But… how should I set $\boldsymbol{w}$?



Deterministic Approach to LR

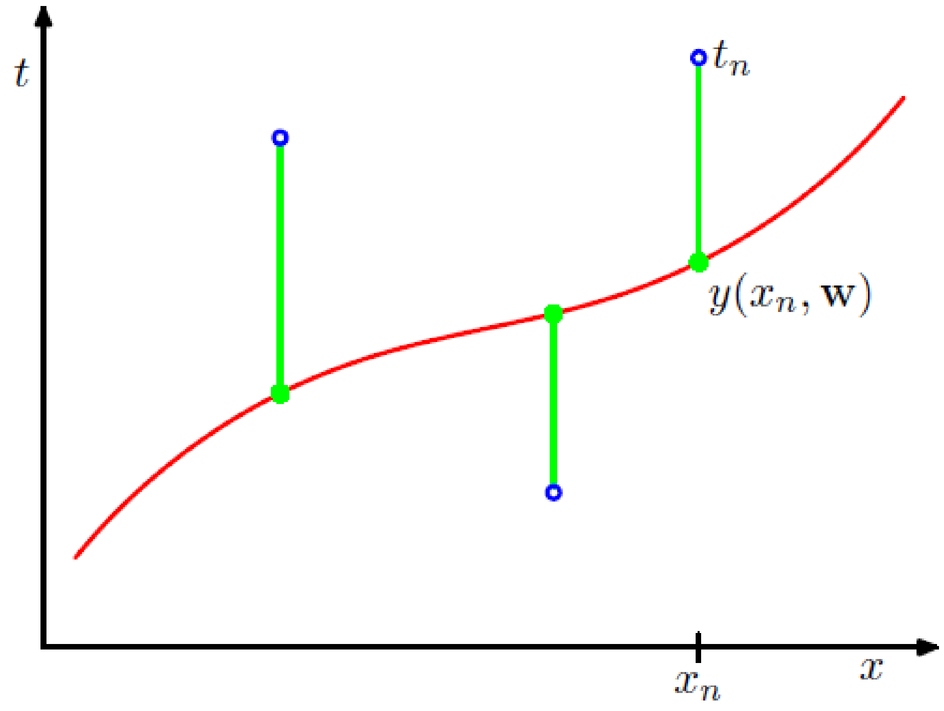
- Data and model parameters considered deterministic

- Quadratic error function

$$E(\boldsymbol{w}) = \sum_{i=1}^{N} \left(y_i - \boldsymbol{w}^t \boldsymbol{\Phi}(x_i)\right)^2$$

- Find $\boldsymbol{w}$ by minimizing L_2 (OLS)

$$\boldsymbol{w} = (\boldsymbol{\Phi}\boldsymbol{\Phi}^t)^{-1}\,\boldsymbol{\Phi}\,\boldsymbol{y}$$



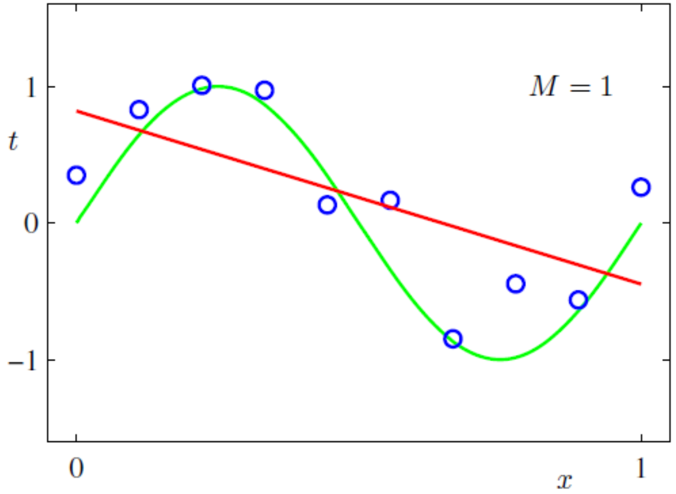
Deterministic Approach to LR

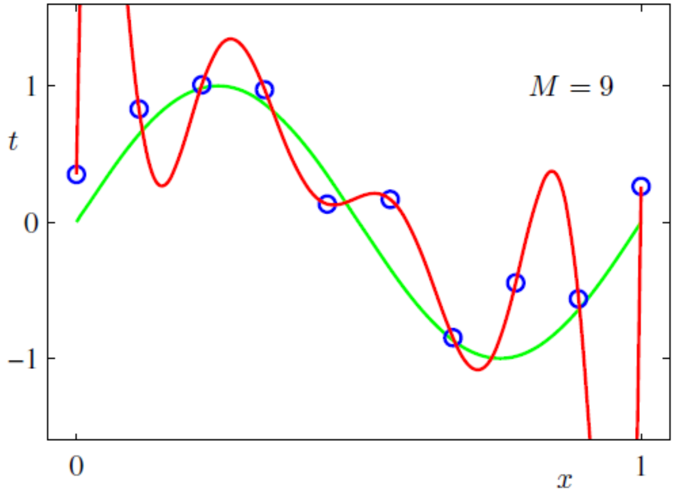
- **Model selection**
  - How complex should the model be?

$$\Phi(x) = [\Phi_1(x), \dots, \Phi_M(x)]$$
$$= [1, x, \dots, x^{M-1}]$$

- Too simple model
  - Poor **fitting**, poor **generalization**

- Too complex model
  - **Overfitting** (very good fitting, poor generalization)

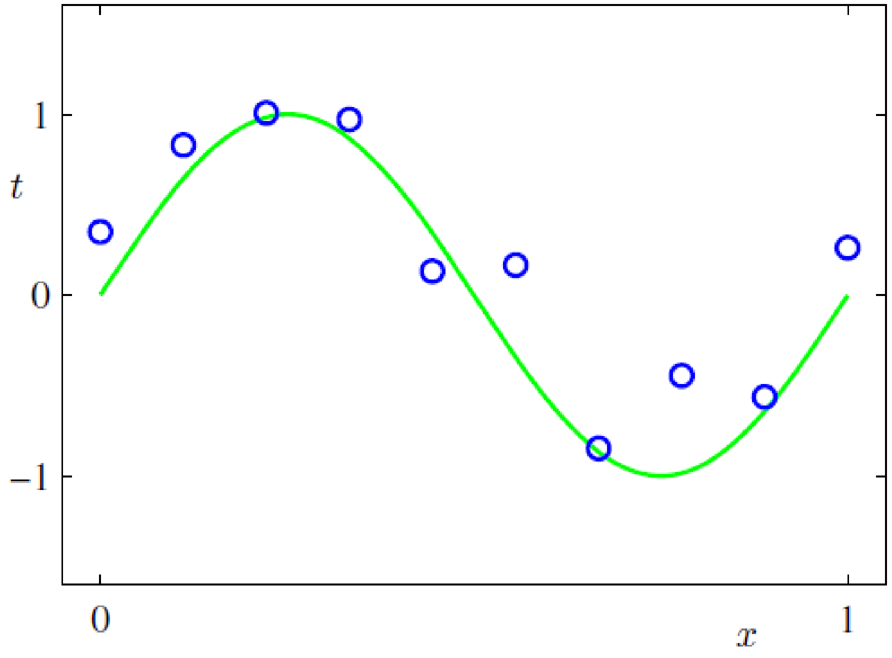




Frequentist Approach to LR

- Data is considered random

$$y_i = m(x_i) + \varepsilon_i, \quad \varepsilon_i \sim N(0, \sigma^2)$$



- Likelihood

$$p(\boldsymbol{y}|\boldsymbol{w}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{\|y_i - m(x_i)\|^2}{2\sigma^2}} = L(\boldsymbol{w}|\boldsymbol{y})$$

- Find model parameters $\boldsymbol{w}$ by maximizing the likelihood

Bayesian Approach to LR

$$p(\boldsymbol{w}|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{w})\, p(\boldsymbol{w})}{p(\boldsymbol{y})}$$

- Prior $p(\boldsymbol{w})$
  - Expert knowledge or naive guess about $\boldsymbol{w}$
- Likelihood $p(\boldsymbol{y}|\boldsymbol{w})$
  - Totally data driven, used to update the prior.
- Posterior $p(\boldsymbol{w}|\boldsymbol{y})$
  - Push and pull game between prior and likelihood

Bayesian Approach to LR

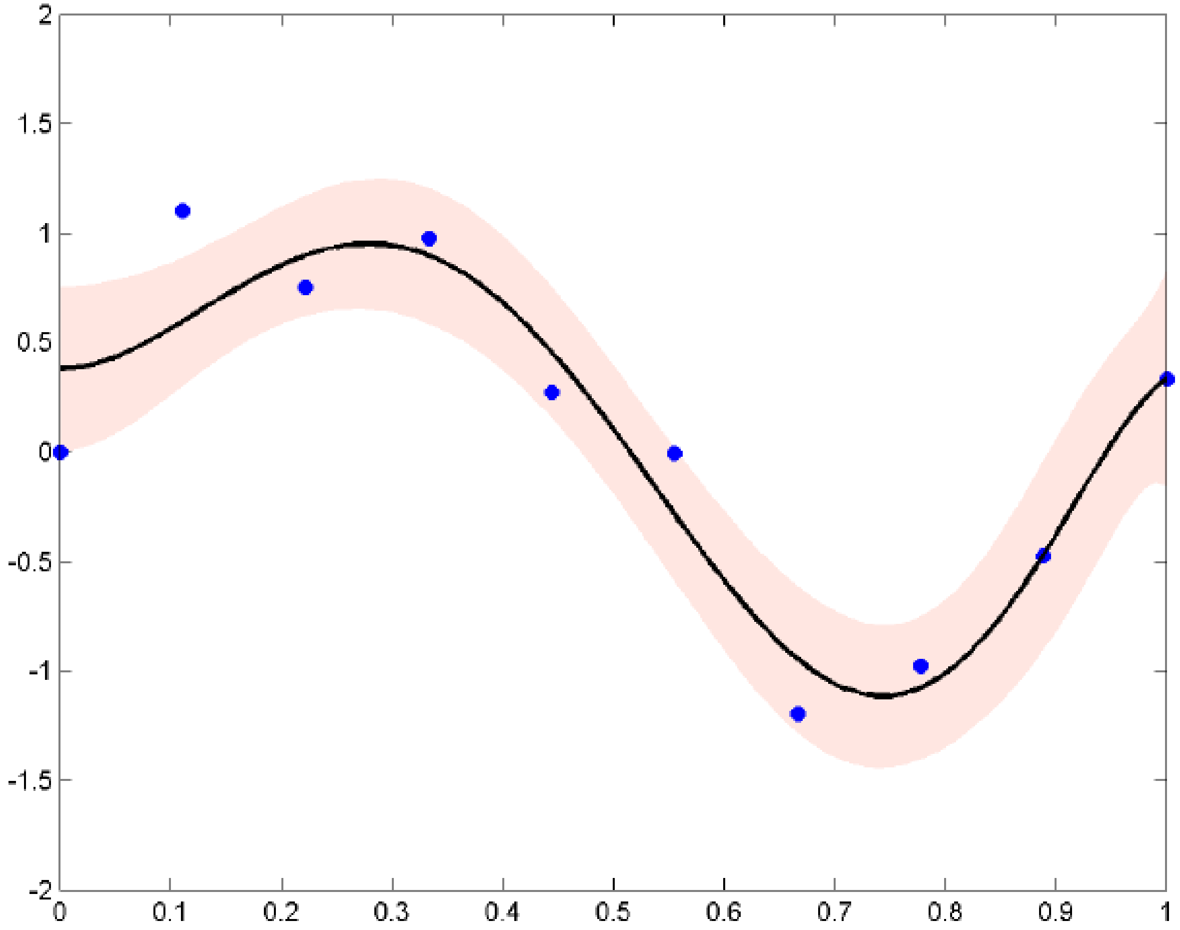
- Predictive distribution

$$p(y^*|\boldsymbol{y}) = \int_{\boldsymbol{w}} p(y^*|\boldsymbol{w})\, p(\boldsymbol{w}|\boldsymbol{y})\, d\boldsymbol{w}$$

- In a fully Gaussian setting

$$p(\boldsymbol{w}|\boldsymbol{y}) \sim N(\widehat{\boldsymbol{w}}, A^{-1})$$

$$p(y^*|\boldsymbol{y}) \sim N\left(\widehat{\boldsymbol{w}}^t \boldsymbol{\Phi}(x^*)\right)$$



Linear Regression Summary

Need to choose a basis function **family** and **number** (not an easy choice)

Deterministic and frequentist approaches

- Focus on fitting the model only based on observations

Bayesian approach

- Allows introduction of prior knowledge
- Sound way for integrating uncertainty
- Inference was made over the model parameters

But… were the model parameters the real unknown?

- Recall our goal: learn an unknown function from data

Bayesian Regression

Based on Gaussian Process for Machine Learning (Rasmussen Williams 2006)

Bayesian Approach to Regression

Bayesian approach:

- All forms of uncertainty are modelled through probabilities

Unknown quantity:

- The function $f(x)$ that we wish to integrate

$$\boxed{I = \int f(x)\,\mathrm{d}x}$$

Solution: use a probabilistic model for the unknown function

- Gaussian Process model
- Characterize our knowledge and certainty about $f(x)$

Gaussian Process

- Rules probabilities over functions

- Formally: a collection of random variables with joint Gaussian distribution
  - Mean function $\bar{f}(x)$
  - Covariance function $k(x, x')$

$$f(x) \sim GP\left(\bar{f}(x), k(x, x')\right),$$

$$\bar{f}(x) = \mathrm{E}[f(x)]$$

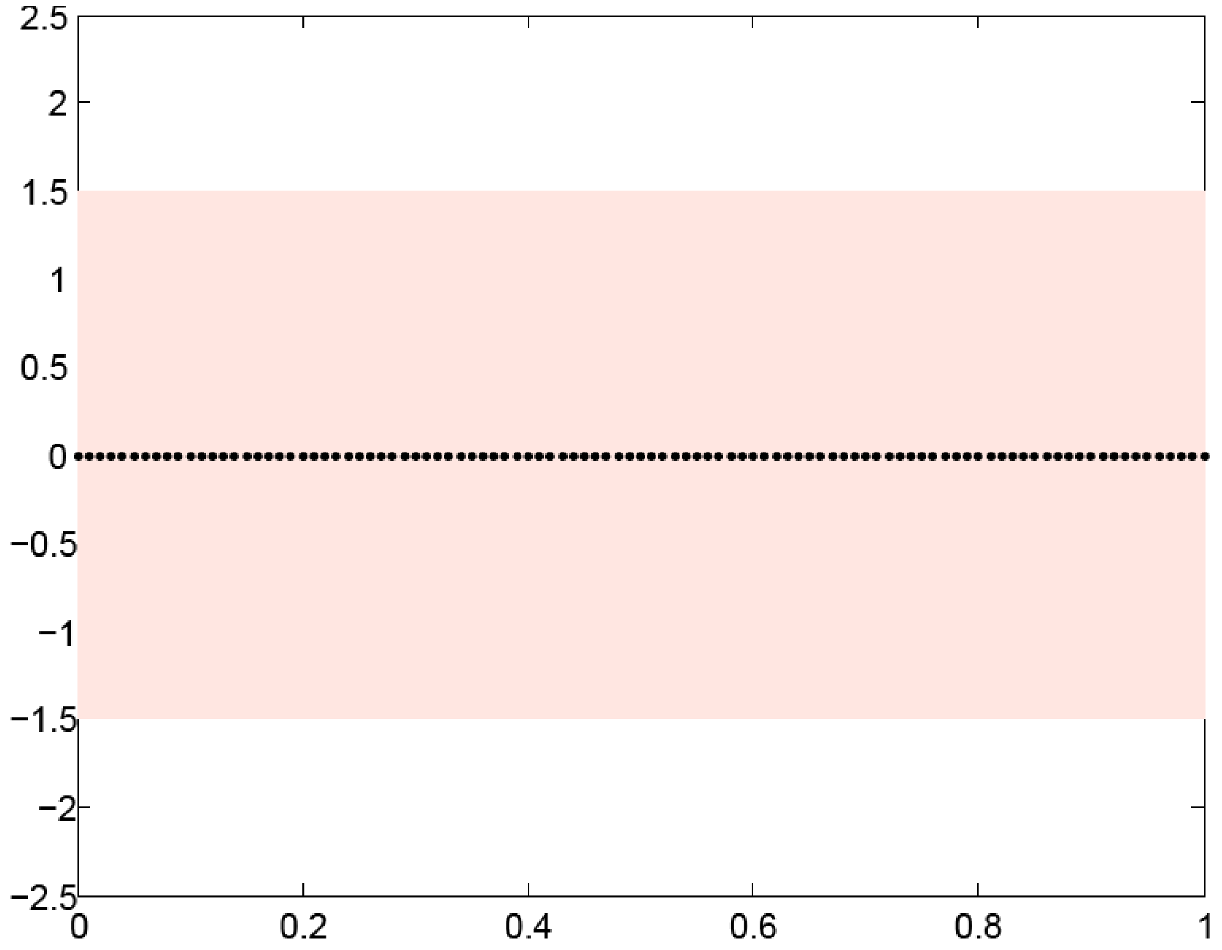
$$k(x, x') = \mathrm{E}\left[\left(f(x) - \bar{f}(x)\right)\left(f(x') - \bar{f}(x')\right)\right]$$

Gaussian Process Prior – Practical Example

- Simple prior over $f(x)$

$$\bar{f}(x) = 0$$

$$k(x, x') = \sigma_f^2 \exp\left(-\frac{1}{2}\frac{\|x - x'\|^2}{l^2}\right)$$



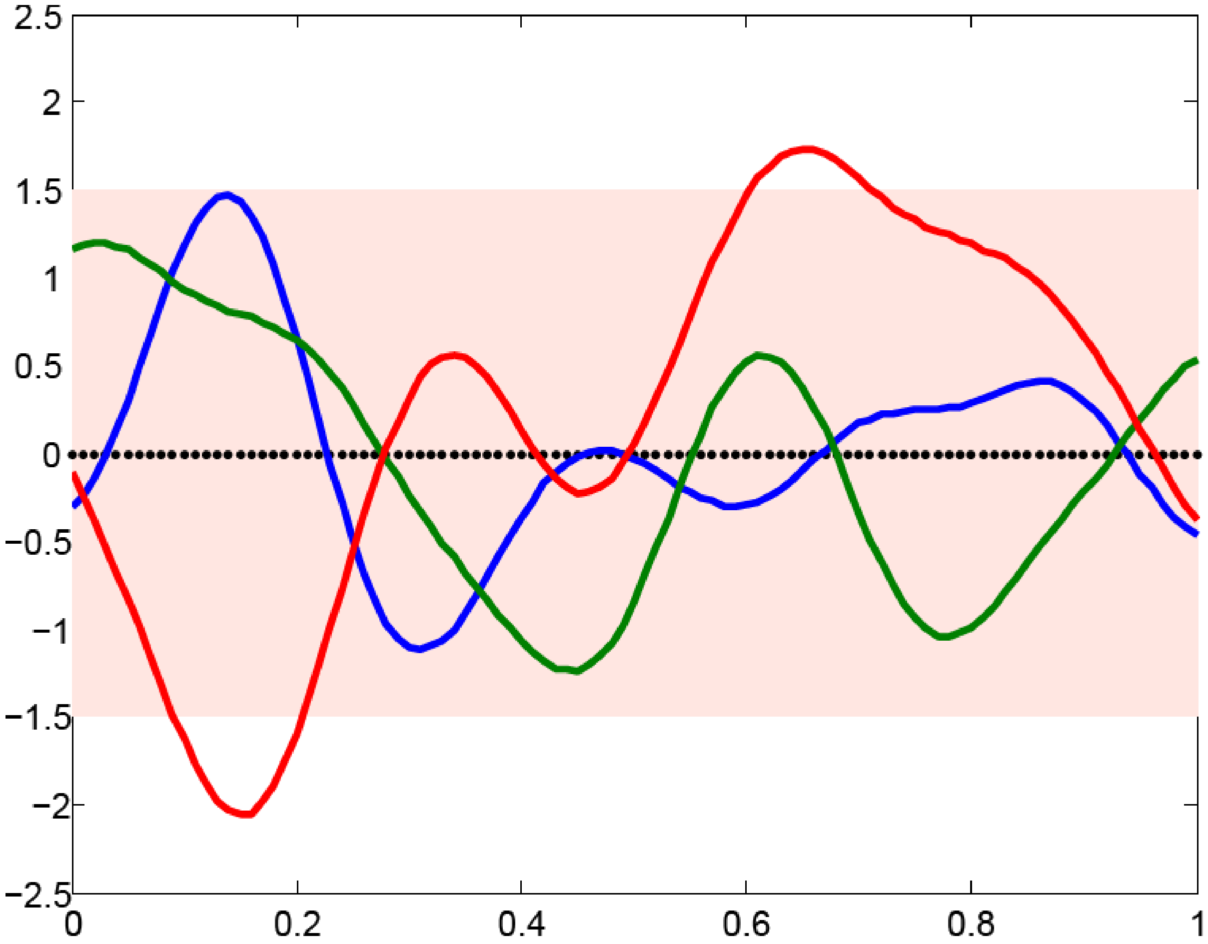
Gaussian Process Prior – Practical Example

- Simple prior over $f(x)$

$$\bar{f}(x) = 0$$

$$k(x, x') = \sigma_f^2 \exp\left(-\frac{1}{2}\frac{\|x - x'\|^2}{l^2}\right)$$

- Labs: assume σ_f and l are known

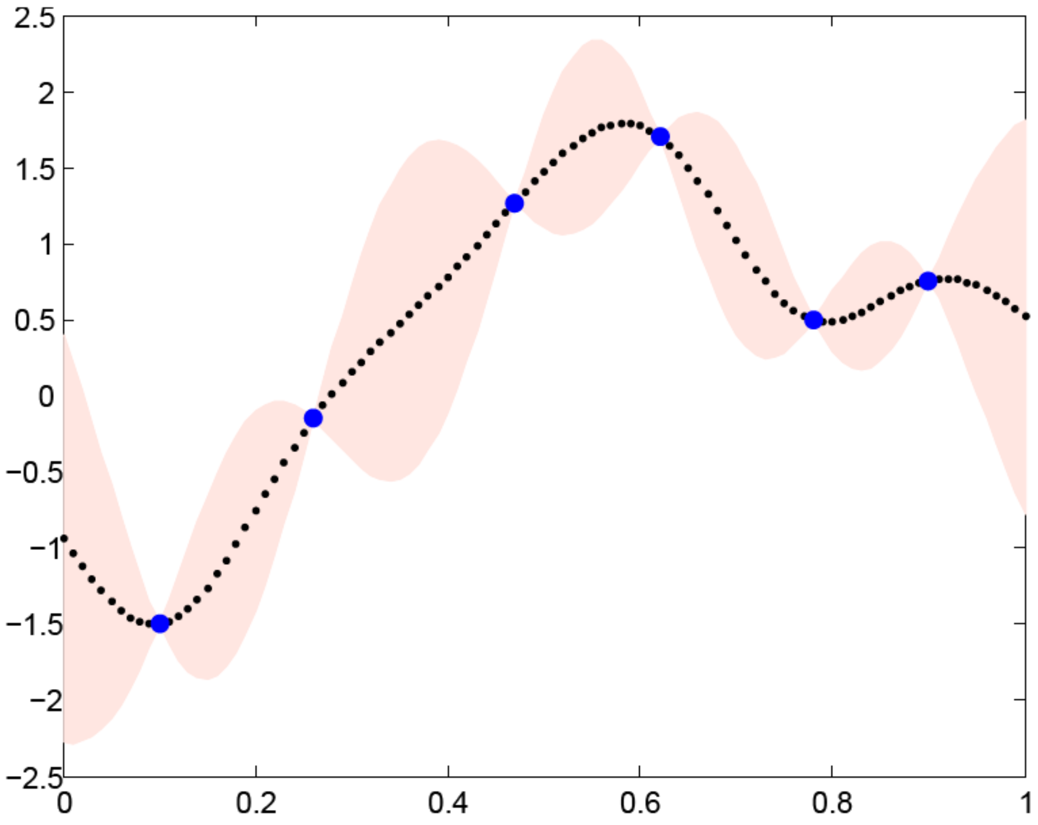


Posterior Gaussian Process

- Set D of N observations

$$D = \{(x_i, y_i) \mid i = 1, \ldots, N\}$$

- Posterior GP
  - Condition prior to D using Bayes' rule
  - Combines prior with evidence
  - Uncertainty (pink) drastically reduced near sample locations

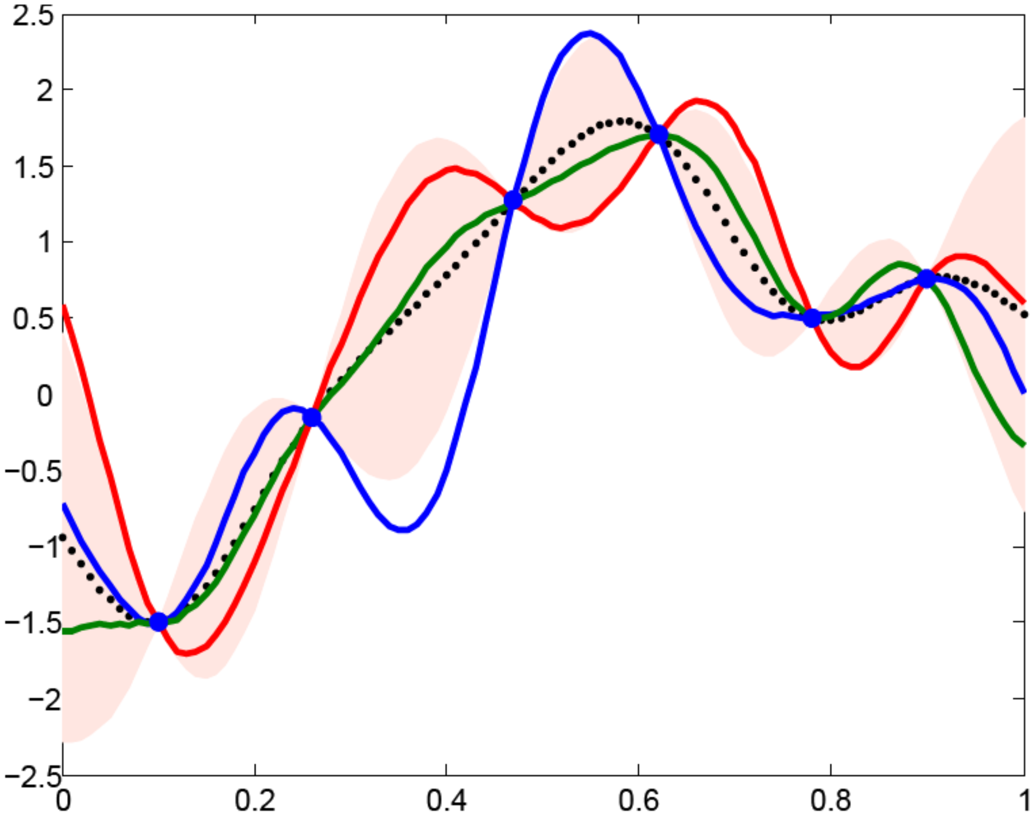


Posterior Gaussian Process

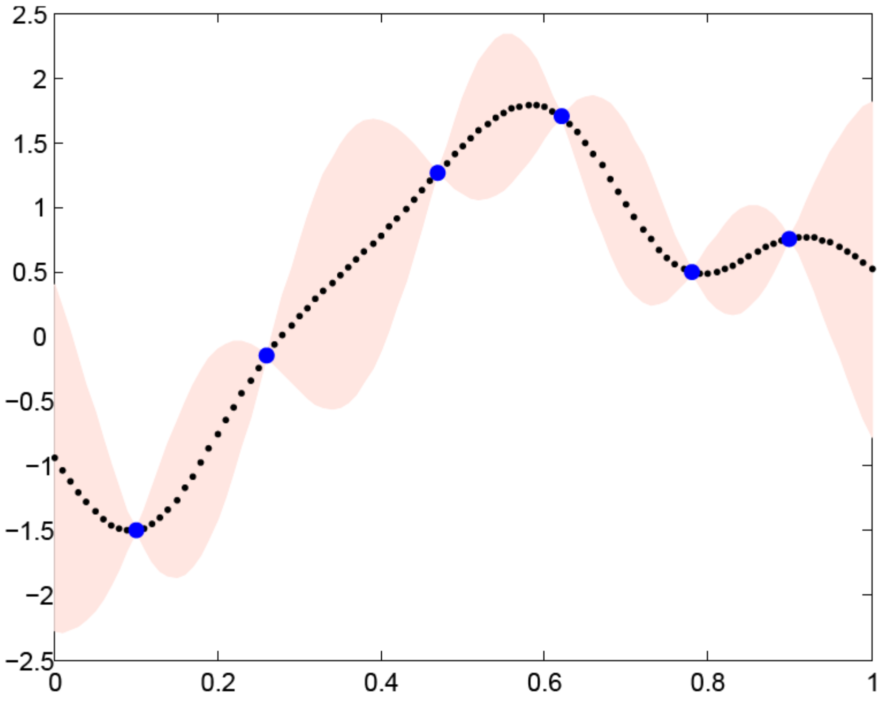
- Set D of N observations

$$D = \{(x_i, y_i) \mid i = 1, \ldots, N\}$$

- Posterior GP
  - Condition prior to D using Bayes' rule
  - Combines prior with evidence
  - Uncertainty (pink) drastically reduced near sample locations



Bayesian Regression Equations



$$E[f(x_*) \mid D] = \bar{f}(x_*) + \boldsymbol{k}(x_*)^t \, Q^{-1} \, (\boldsymbol{y} - \overline{\boldsymbol{y}})$$

$\bar{f}(x)$ is the prior expected value of the unknown function $f(x)$

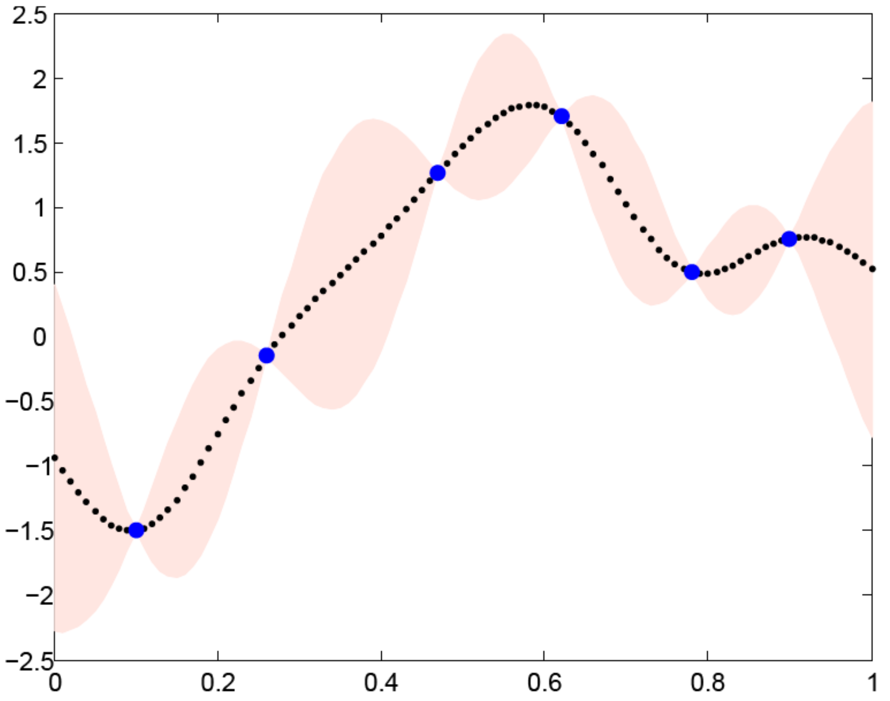
$\boldsymbol{k}(x)$ is a vector which captures the relevance of the sample position

Q is a matrix which captures the inter-samples distance

$\boldsymbol{y}$ is a vector of observations

$\overline{\boldsymbol{y}}$ is the prior expected value of the observations $\boldsymbol{y}$

Bayesian Regression Equations



$$E[f(x_*)\mid D] = \bar{f}(x_*) + \boldsymbol{k}(x_*)^t\, Q^{-1}\, (\boldsymbol{y} - \bar{\boldsymbol{y}})$$

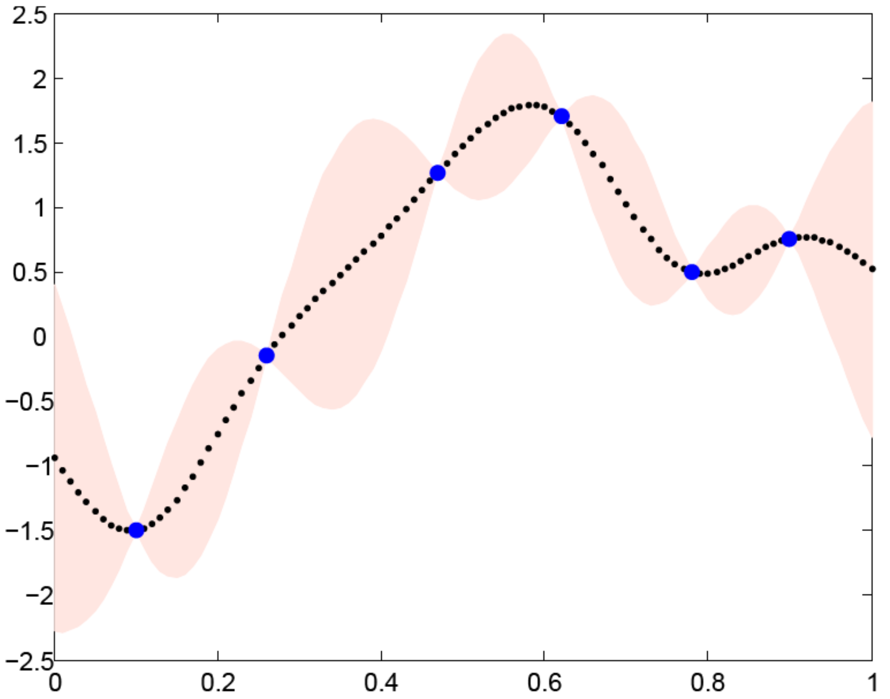
$$\boldsymbol{k}(x) = \big(k(x_1, x_*), \dots, k(x_N, x_*)\big)$$

$$Q = \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_N) \\ \vdots & \ddots & \vdots \\ k(x_N, x_1) & \cdots & k(x_N, x_N) \end{bmatrix}$$

$$\boldsymbol{y} = \big(f(x_1), \dots, f(y_N)\big)^t$$

$$\bar{\boldsymbol{y}} = \left(\bar{f}(x_1), \dots, \bar{f}(x_N)\right)^t$$

Bayesian Regression Equations



$$\text{cov}[f(x), f(x') \mid D] = k(x, x') - \boldsymbol{k}(x)^t \, Q^{-1} \, \boldsymbol{k}(x)$$

$$\boldsymbol{k}(x) = \left[k(x_1, x_*), \ldots, k(x_N, x_*)\right]$$

$$Q = \begin{bmatrix} k(x_1, x_1) & \cdots & k(x_1, x_N) \\ \vdots & \ddots & \vdots \\ k(x_N, x_1) & \cdots & k(x_N, x_N) \end{bmatrix}$$

GP Live Demo

http://chifeng.scripts.mit.edu/stuff/gp-demo/

Interesting 20 min video on Gaussian Processes

- Link here

Bayesian Monte Carlo

Based on the “Bayesian Monte Carlo” paper of *Ghahramani and Rasmussen, 2006*

Bayesian Monte Carlo – Problem Setting

Consider an integral of the form

$$I = \int f(x)\, p(x)\, \mathrm{d}x$$

where

- $f(x)$ is unknown
- $p(x)$ is known

Most common case in rendering problems:

- $f(x)$ is the incident radiance function $L_i(\omega_i)$
- $p(x)$ is the product between the BRDF and the cosine term

Bayesian Monte Carlo - Approach

State the prior knowledge about $f(x)$ in a probabilistic fashion

- Use a GP model of $f(x)$

Refine the GP model using a set of observations (samples)

- Yields a posterior GP model of $f(x)$

The Bayesian integral estimate is then given by:

$$\hat{I}_{BMC} = E[I \mid D] = \int \tilde{f}(x)\, p(x)\, \mathrm{d}x$$

$\tilde{f}(x)$ is the most probable value of $f(x)$ according to the posterior GP

Bayesian Monte Carlo Equations

The BMC equations: integrate both sides of the BR equations

$$\hat{I}_{BMC} = E[I \mid D] = \bar{I} + z^t Q^{-1} \left(\boldsymbol{Y} - \overline{\boldsymbol{Y}}\right)$$

$\bar{I} = \int \bar{f}(x)\, p(x)\, dx$

$\boldsymbol{Y} = (y_1, \dots, y_N)^t$

Q is the covariance matrix

$\overline{\boldsymbol{Y}} = \left(\bar{f}(x_1), \dots, \bar{f}(x_N)\right)^t$

$$\mathrm{z} = \left(\int k(x_1, x)\, p(x)\, dx, \dots, \int k(x_N, x)\, p(x)\, dx\right)$$

Bibliography

- “Pattern Recognition and Machine Learning”, C. Bishop (2006)
- “Gaussian Process for Machine Learning”, C. Rasmussen & C. Williams (2006)
- “Bayesian Monte Carlo”, Rasmussem and Ghahramani (2006)

Hands on Code

Practice 3 (BMC)