

Gaussian Processes For Machine Learning

2025-12-07

Table of contents

| | | |
|---|--------------------------|---|
| 1 | Preface | 3 |
| 2 | Chapter 01: Introduction | 5 |

1 Preface

[R] Start around 7:15PM EST, 2026-01-11.

[A] BOOK for Gaussian Processes For Machine Learning (2006) by Carl Edward Rasmussen and Christopher K. I. Williams as Rasmussen_GPFML_2006.

[>] The actual science of logic is conversant at present only with things either certain, impossible, or entirely doubtful, none of which (fortunately) we have to reason on. Therefore the true logic for this world is the calculus of Probabilities, which takes account of the magnitude of the probability which is, or ought to be, in a reasonable man's mind. — James Clerk Maxwell [1850].

[>] "...building systems that can adapt to their environment and learn from their experience..." (Series Forward)

[A] GP for Gaussian Process.

[R] Some high-level descriptions of the scope of the book include: "...the mathematical foundations and practical application of Gaussian processes in regression and classification tasks" and "...how Gaussian processes can be interpreted as a Bayesian version of the well-known support vector machine methods." (Series Forward).

[R] The goal of the book: "...provide a systematic and unified treatment of this area [”the application of Gaussian process models to machine learning task”].”

[>] “Gaussian processes provide a principled, practical, probabilistic approach to learning in kernel machines.”

[>] “Roughly speaking a stochastic process is a generalization of a probability Gaussian process distribution (which describes a finite-dimensional random variable) to functions.”

[T] Define stochastic process, probability distribution, Gaussian, kernel, support vector machine, random variable, function, basis function, supervised learning, unsupervised learning, ...

[>] “By focussing on processes which are Gaussian, it turns out that the computations required for inference and learning become relatively easy.”

[>] “While his own work was focused on sophisticated Markov chain methods for inference in large finite networks, he did point out that some of his networks became Gaussian processes in the limit of infinite size”

[A] NN for Neural Network.

[R] You appreciate the historical remarks concerning NNs and GPs made by the authors.

[E] What did NNs become so popular?

...they allowed the use of adaptive basis functions, as opposed to the well known linear models.

[R] There are likely some very interesting retrospective analyses that can be done with respect to claims and statements made in this Preface concerning aspects of the future of the machine learning (written before the revolution in deep learning).

[T] The preface ought to be read again after the rest of the book is read.

[Q] What is a “more expressive” covariance function?

[E] How do the goals of statistics and machine learning differ?

...in statistics a prime focus is often in data and models understanding the data and relationships in terms of models giving approximate summaries such as linear relations or independencies. In contrast, the goals in machine learning are primarily to make predictions as accurately as possible and algorithms and predictions to understand the behaviour of learning algorithms.

[E] How do Gaussian processes bridge the gap between machine learning and statistics?

Gaussian processes are mathematically equivalent to many well known models, including Bayesian linear models, spline models, large neural networks (under suitable conditions), and are closely related to others, such as support vector machines. Under the Gaussian process viewpoint, the models may be easier to handle and interpret than their conventional counterparts, such as e.g. neural networks.

[T] Parse through the mathematical abbreviations and ensure you understand the different terms presented there.

2 Chapter 01: Introduction

[R] Start at 10:06 PM EST, 2026-01-11.

[T] Define the terms: regression, classification, ...

[T] Come up with 5 examples of classification problems and 5 examples of regression problems?

[I] An interesting blog-post might be recording different dataset sizes of your own hand-written digits or letters and then using different methods for classifying hand-written digits or letters and seeing how each performs relative to the other with different amounts of training data.

[R] How to write, mathematically, a dataset: \mathcal{D} of n observations, $\mathcal{D} = \{(\mathbf{x}_i, y_i) | i = 1, \dots, n\}$ where \mathbf{x} is the input and y is the output (or target).

[R] “training is inductive” as “we need to move from the finite training data \mathcal{D} to a function f that makes predictions for all possible input values.”

[R] Two common approaches for dealing with the supervised learning problem (and the “assumptions about the characteristics of the underlying function”):

[R] End at 10:15 PM EST, 2026-01-11.