Boosting Derivatives Pricing with Machine Learning

Jan De Spiegeleer

Risk Analytics Symposium 2019
Chicago

Joint work with Dilip Madan, Wim Schoutens, Sofie Reyners
Content

• Machine Learning for Quant Problems
• Gauss Process Regression Explained
• Pricing of Exotics
• Comparing GPR with GBM
• Conclusions
Introduction
Derivative pricing and risk management is time-consuming

- Exotic option pricing
  → *Monte Carlo simulations*
- VaR and ES
  → *Scenarios*
- Counter-party exposure
... and markets are moving!

time-consuming algorithms

continuously moving markets

→ prices are outdated when available, overnight calculations cannot be performed in one night, ...
Machine Learning techniques can help here.

When training is completed, prediction is extremely fast!
Gaussian Process Regression
Assume $f(x)$ to be multi-variate Normal (GP)
“Forcing” a Structure upon the GP

\[ f \sim N(0, I) \quad \text{and} \quad f \sim N(0, \Sigma) \]
We have the function value 2 points $x_1 = 20$ and $x_2 = 40$. The corresponding values are $f_1 = 0$ and $f_2 = 1$. What is the value in point $x_3$?

The only thing we have are the data-points ($f_1$ and $f_2$) and the fact that $f$ is a Gaussian process.

The only freedom we have is the choice of the Kernel function $K()$ and its hyper-parameters.
Using Kernels

The kernel $K(a, b)$ of two points $a$ and $b$ gives a measure of the "distance" of these points.

Examples of kernel functions

- **Radial Basis Function** $K(a, b) = \exp\left(\frac{1}{\sigma^2} \| a - b \|^2\right)$

- **Matern** $K(a, b) = \sigma^2 \frac{1}{\Gamma(\nu)\left(2\nu\right)^{\nu}} \left(\gamma\sqrt{2\nu d(a/l, b/l)}\right)\nu K_\nu\left(\gamma\sqrt{2\nu d(a/l, b/l)}\right)$

- **RationalQuadratic** $K(a, b) = \left(1 + \frac{d(a,b)^2}{2\alpha l^2}\right)^{-\alpha}$

- **ExpSineSquared** $K(a, b) = \exp\left(-2 \left(\sin(\pi/p * d(a, b)/l)\right)^2\right)$

Reference:
Carl Eduard Rasmussen and Christopher K.I. Williams, “Gaussian Processes for Machine Learning”, MIT Press 2006,
Model = Data

\[
\begin{bmatrix}
  f_1 \\
  f_2 \\
  f_3
\end{bmatrix}
\sim N \left( \begin{bmatrix}
  0 \\
  0 \\
  0
\end{bmatrix}, \begin{bmatrix}
  k_{11} & k_{12} & k_{13} \\
  k_{21} & k_{22} & k_{23} \\
  k_{31} & k_{32} & k_{33}
\end{bmatrix} \right)
\]
\[
\begin{bmatrix}
  f_1 \\
  f_2 \\
  f_3 
\end{bmatrix}
\sim N\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} k_{11} & k_{12} & k_{13} \\ k_{21} & k_{22} & k_{23} \\ k_{31} & k_{32} & k_{33} \end{bmatrix}\right)
\]

\[
\begin{bmatrix}
  f \\
  f_* 
\end{bmatrix}
\sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} K & K_* \\ K^T & K^{**} \end{bmatrix}\right)
\]

Function value at \(x=30\)?
\[
\begin{pmatrix} f & f_* \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix} , \begin{pmatrix} K & K_* \\ K_*^T & K** \end{pmatrix} \right)
\]

Function value at \(x=30\): 

\(f_* |_{x,x';f} \sim N(\mu, \Sigma)\)

\[
\mu = K_*^T K^{-1} f
\]
\[
\Sigma = K** - K_*^T K^{-1} K_*
\]
Back to introductory example (1)

Using the following Kernel: \( k(a, b) : \exp \left( -\frac{1}{2} \frac{(a-b)^2}{l} \right) \) with \( l = 100 \)

\[
\begin{array}{|c|c|c|}
\hline
x & 20.0 & 40.0 & 30.0 \\
\hline
20.0 & 1.00 & 0.14 & 0.61 \\
40.0 & 0.14 & 1.00 & 0.61 \\
30.0 & 0.61 & 0.61 & 1.00 \\
\hline
\end{array}
\]

\[
\begin{align*}
K(x = 20.0, x = 40.0) &= 1.00 \\
K(x = 40.0, x = 20.0) &= 0.14 \\
K(x = 30.0, x = 40.0) &= 0.61 \\
K(x = 40.0, x = 30.0) &= 1.00 \\

K_{**}(x = 30.0, x = 30.0) &= 1.00 \\

\mu &= K^T K^{-1} f \\
\Sigma &= K_{**} - K^T K^{-1} K \\

f_*(x = 30) &\sim N(0.51, 0.35) \\
\mu &= 0.51 \text{ and } \Sigma = 0.35
\end{align*}
\]
Back to introductory example (2)

$$f_\star = f(x = 30) \sim N(0.51, 0.35)$$
More points to predict

<table>
<thead>
<tr>
<th></th>
<th>$x = 20.0$</th>
<th>$x = 30.0$</th>
<th>$x = 40.0$</th>
<th>$x = 50.0$</th>
<th>$x = 60.0$</th>
<th>$x = 70.0$</th>
<th>$x = 80.0$</th>
<th>$x = 90.0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x = 20.0$</td>
<td>1.00</td>
<td>0.61</td>
<td>0.14</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$x = 30.0$</td>
<td>0.61</td>
<td>1.00</td>
<td>0.61</td>
<td>0.14</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$x = 40.0$</td>
<td>0.14</td>
<td>0.61</td>
<td>1.00</td>
<td>0.61</td>
<td>0.14</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$x = 50.0$</td>
<td>0.01</td>
<td>0.14</td>
<td>0.61</td>
<td>1.00</td>
<td>0.61</td>
<td>0.14</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>$x = 60.0$</td>
<td>0.00</td>
<td>0.01</td>
<td>0.14</td>
<td>0.61</td>
<td>1.00</td>
<td>0.61</td>
<td>0.14</td>
<td>0.01</td>
</tr>
<tr>
<td>$x = 70.0$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.14</td>
<td>0.61</td>
<td>1.00</td>
<td>0.61</td>
<td>0.14</td>
</tr>
<tr>
<td>$x = 80.0$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.14</td>
<td>0.61</td>
<td>1.00</td>
<td>0.61</td>
</tr>
<tr>
<td>$x = 90.0$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.14</td>
<td>0.61</td>
<td>1.00</td>
</tr>
</tbody>
</table>

$K$

<table>
<thead>
<tr>
<th></th>
<th>$x = 20.0$</th>
<th>$x = 40.0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x = 20.0$</td>
<td>1.00</td>
<td>0.61</td>
</tr>
<tr>
<td>$x = 40.0$</td>
<td>0.61</td>
<td>1.00</td>
</tr>
</tbody>
</table>

$K_*$

<table>
<thead>
<tr>
<th></th>
<th>$x = 30.0$</th>
<th>$x = 50.0$</th>
<th>$x = 60.0$</th>
<th>$x = 70.0$</th>
<th>$x = 80.0$</th>
<th>$x = 90.0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x = 20.0$</td>
<td>0.14</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$x = 40.0$</td>
<td>0.61</td>
<td>0.14</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

$K_{**}$

<table>
<thead>
<tr>
<th></th>
<th>$x = 30.0$</th>
<th>$x = 50.0$</th>
<th>$x = 60.0$</th>
<th>$x = 70.0$</th>
<th>$x = 80.0$</th>
<th>$x = 90.0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x = 30.0$</td>
<td>1.00</td>
<td>0.61</td>
<td>0.14</td>
<td>0.01</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>$x = 50.0$</td>
<td>0.61</td>
<td>1.00</td>
<td>0.61</td>
<td>0.14</td>
<td>0.01</td>
<td>0.00</td>
</tr>
<tr>
<td>$x = 60.0$</td>
<td>0.14</td>
<td>0.61</td>
<td>1.00</td>
<td>0.61</td>
<td>0.14</td>
<td>0.01</td>
</tr>
<tr>
<td>$x = 70.0$</td>
<td>0.01</td>
<td>0.14</td>
<td>0.61</td>
<td>1.00</td>
<td>0.61</td>
<td>0.14</td>
</tr>
<tr>
<td>$x = 80.0$</td>
<td>0.00</td>
<td>0.01</td>
<td>0.14</td>
<td>0.61</td>
<td>1.00</td>
<td>0.61</td>
</tr>
<tr>
<td>$x = 90.0$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.01</td>
<td>0.14</td>
<td>0.61</td>
<td>1.00</td>
</tr>
</tbody>
</table>
More points to predict

\[ f^* |_{x^*, x, f} \sim N(\mu, \Sigma) \]
More points to predict
More points to predict + Extra point in training set

\[ f(x = 80) = -1 \]
Where is the caveat?

\[
\begin{align*}
\mu &= K^T K^{-1} f \\
\Sigma &= K_{**} - K^T K^{-1} K_*
\end{align*}
\]

• The size of the matrix K grows with the training set
• Taking the inverse can be a computational challenge
• GPR is a technological challenge.
GPR in Derivatives Pricing
GPR Idea

![Analytical Model](Fig 1.1)

![Grid Solution](Fig 1.2)

Using Basis Function $\Phi(x) = [1, s, s^2, s^3]$

![Grid-Interpolation](Fig 1.3)
GPR Idea

• The GPR approach is a *non-parametric* approach, in that it finds a distribution over the possible *functions* $f(x)$ that are consistent with the observed data.

• As with all Bayesian methods it begins with a prior distribution and updates this as data points are observed, producing the posterior distribution over functions.
Not too wiggly

• *Not too wiggly*... what does that mean?
• Here on the left is an example of a very wiggly function. And at the right here’s a much smoother function:

![](image1.png) ![](image2.png)

• We use a **covariance matrix** to ensure that values that are close together in input space will produce output values that are close together. Covariance is a way to specify that smoothness:

  low covariance = no relation  \[\text{high covariance} = \text{similar / close together}\]

• This covariance matrix, along with a mean function to output the expected value of \(f(x)\) defines a Gaussian Process.
GPR Definition

Consider a training set

\[(X, y) = \{(x_i, y_i) \mid i = 1, \ldots, n\}\].

Find a relation between inputs and outputs:

\[y_i = f(x_i) + \epsilon_i\]

where \(f(x)\) is a Gaussian process and \(\epsilon_i \sim \mathcal{N}(0, \sigma_n^2)\) are i.i.d. random variables representing the noise in the data.
GP: a Bayesian method

Posterior distribution

Only consider functions that agree with the data.

- Take new inputs $X_*$, with corresponding (unknown) function values $f_*$

- Joint distribution of training outputs and function values:

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim \mathcal{N} \left( 0, \begin{bmatrix} K(X, X) + \sigma_n^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix} \right)$$
GP: a Bayesian method

**Posterior distribution**

- Condition on the observations:

\[
\begin{align*}
  f_*|X_*, X, y & \sim \mathcal{N}(\mu, \Sigma) \\
  \mu & = K(X_*, X) [K(X, X) + \sigma_n^2 I]^{-1} y \\
  \Sigma & = K(X_*, X_*) - K(X_*, X) [K(X, X) + \sigma_n^2 I]^{-1} K(X, X_*)
\end{align*}
\]
GP Kernels

- Kernel functions: several choices are possible, each with different hyperparameters.

- How to determine the hyperparameters?

Maximize the marginal (log-)likelihood:

$$-rac{1}{2} \log(\det(K)) - \frac{1}{2} y^T K^{-1} y + \text{const}$$
Barrier Option Pricing

Pricing of Down-and-Out-Barrier Put (T=1, H=75, K=140, \( \theta = 100 \)) - Learning on 5 Points
Barrier Option Pricing
Barrier Option Pricing
Structured Products

Model: Heston

Training set:

- \( B : 1.05 \rightarrow 1.55 \)
- \( H : 0.55 \rightarrow 0.95 \)
- \( q : 0 \rightarrow 0.05 \)
- \( r : 0.02 \rightarrow 0.03 \)
- \( T : 11/12 \rightarrow 1 \)
- \( \rho : -0.95 \rightarrow -0.25 \)
- \( v_0 : 0.01 \rightarrow 0.25 \)
- \( \kappa : 0.2 \rightarrow 1.6 \)
- \( \theta : 0.15 \rightarrow 0.65 \)
- \( \epsilon : 0.01 \rightarrow 0.25 \)

Assume \( S_0 = 1 \).

At expiry a holder receives

A. the final index level if it is above 140% of the initial level, or
B. 140% of the initial level, if the final level is between 75% and 140% of the initial level, unless the index level has fallen below 75% of the initial level during the lifetime of the certificate in which case...
C. one receives just the final level
Structured Products

Out-of-sample GPR prediction - ARD SE kernel, zero mean

BC (GPR)

BC (MC)
Performance

GPR Performance

• Max. absolute error: 0.0057
• Mean absolute error: 0.0006
• Max. relative error: 0.0045
• Mean relative error: 0.0005

Speed-up: $\times 600$

$n_{\text{train}} = 10\,000$
$n_{\text{test}} = 100\,000$
GPR: built-in Greeks

Conclusion and more info

• We have presented new Machine Learning tools for solving traditional quant problems.
• Key is a fast matrix inversion of a huge kernel matrix.
• Speed-ups of several magnitudes can be achieved once the model is trained.

RiskConcile Solution: Plug-and-Price
Contact: wim@schoutens.be, jds@riskconcile.com

Reference