Гауссовские процессы. Часть 2

Максим Кочуров

МГУ им. М.В. Ломоносова

Лекция 5

Содержание

Временные ряды, классический подход

Если в данных есть сезонность, обычно используют STL декомпозицию (“**S**easonal and **T**rend decomposition using **L**oess”). Однако:

- Параметры не интерпретируемы, есть только декомпозиция
- Нет оценок неопределённости
- Недостаточная гибкость

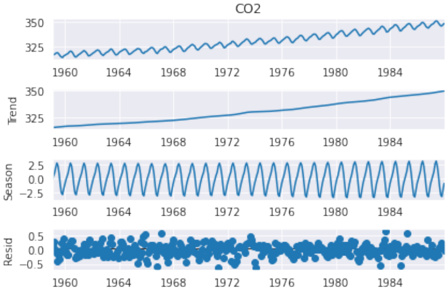


Figure: STL декомпозиция для данных об уровне CO2, Statsmodels

Декомпозиция с гауссовскими процессами

Гауссовские процессы поддерживают дополнительные по сравнению с STL предположения

- Детализация сезонности (год + квартал + месяц + неделя)
- Модели с точками излома (changepoint).
- Гибкая функция правдоподобия
- Модели панельной регрессии
- Пропущенные значения

Типичная модель

Типичная модель аддитивная

$$x_t \sim \underbrace{g(t)}_{\text{тренд}} + \underbrace{s(t)}_{\text{периодичность}} + \underbrace{h(t)}_{\text{праздники}}$$

Источник

См. препринт о моделях Prophet [2]. Каждая модель временного ряда уникальна.

Напоминание

$x \in \mathbb{R}^n$, $y \in \mathbb{R}$

$$Y \sim \mathcal{GP}(m(x), k(x, x'))$$

Напоминание

$x \in \mathbb{R}^n$, $y \in \mathbb{R}$

$$Y \sim \mathcal{GP}(m(x), k(x, x'))$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(x_1) \\ \vdots \\ m(x_N) \end{bmatrix}, \begin{bmatrix} k(x_1, x_1) & \dots & k(x_1, x_N) \\ \vdots & \ddots & \vdots \\ k(x_N, x_1) & \dots & k(x_N, x_N) \end{bmatrix} \right)$$

1. $\mathcal{GP}$ гауссовский процесс – многомерное нормальное распределение с матожиданием $m(t)$ и ковариацией $k(x, x')$

Напоминание

$x \in \mathbb{R}^n$, $y \in \mathbb{R}$

$$Y \sim \mathcal{GP}(m(x), k(x, x'))$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(x_1) \\ \vdots \\ m(x_N) \end{bmatrix}, \begin{bmatrix} k(x_1, x_1) & \dots & k(x_1, x_N) \\ \vdots & \ddots & \vdots \\ k(x_N, x_1) & \dots & k(x_N, x_N) \end{bmatrix} \right)$$

1. $\mathcal{GP}$ гауссовский процесс – многомерное нормальное распределение с матожиданием $m(t)$ и ковариацией $k(x, x')$
2. $m(x)$ – функция среднего
   - Линейная регрессия: $m(x) = x^\top \beta$
   - Константа (в т.ч. ноль): $m(x) = c$
   - Произвольная функция: $m(x) = \sin(x)$

Напоминание

$x \in \mathbb{R}^n$, $y \in \mathbb{R}$

$$Y \sim \mathcal{GP}(m(x), k(x, x'))$$

$$\begin{bmatrix} y_1 \\ \vdots \\ y_N \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} m(x_1) \\ \vdots \\ m(x_N) \end{bmatrix}, \begin{bmatrix} k(x_1, x_1) & \dots & k(x_1, x_N) \\ \vdots & \ddots & \vdots \\ k(x_N, x_1) & \dots & k(x_N, x_N) \end{bmatrix} \right)$$

1. $\mathcal{GP}$ гауссовский процесс – многомерное нормальное распределение с матожиданием $m(t)$ и ковариацией $k(x, x')$
2. $m(x)$ – функция среднего
   - Линейная регрессия: $m(x) = x^\top \beta$
   - Константа (в т.ч. ноль): $m(x) = c$
   - Произвольная функция: $m(x) = \sin(x)$
3. $k(x, x')$ – ядерная функция, мера сходства между x и x'
   - $[K]_{ij} = k(x_i, x_j)$ симметричная положительно определённая матрица

Непериодическая часть (среднее)

- Модели роста
- Модели линейного тренда
- Модели с точками излома (changepoint)

Непериодическая часть (среднее)

- Модели роста
- Модели линейного тренда
- Модели с точками излома (changepoint)

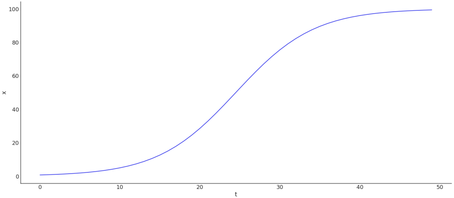


Figure: Модель роста

$$x = \frac{c}{1 + \exp(-k(t - m))}$$

Непериодическая часть (среднее)

- Модели роста
- Модели линейного тренда
- Модели с точками излома (changepoint)

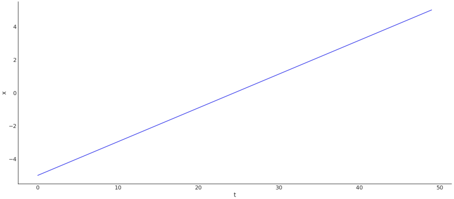


Figure: Модель линейного тренда

$$x = \frac{c}{1 + \exp(-k(t - m))}$$

Непериодическая часть (среднее)

- Модели роста
- Модели линейного тренда
- Модели с точками излома (changepoint)

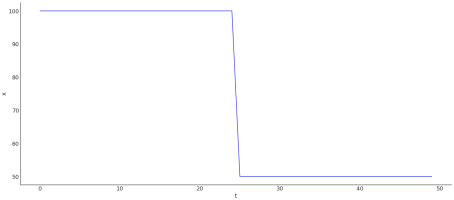


Figure: Модель с точкой излома

$$x = \begin{cases} c_1, & t < m \\ c_2, & t \geq m \end{cases}$$

Непериодическая часть (среднее)

- Модели роста
- Модели линейного тренда
- Модели с точками излома (changepoint)

Расширения

Возможны расширения, например, зависящее от времени насыщение в модели роста, см. [2].

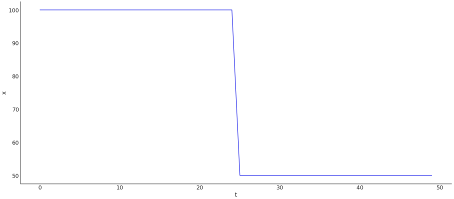


Figure: Модель с точкой излома

$$x = \begin{cases} c_1, & t < m \\ c_2, & t \geq m \end{cases}$$

Праздники

$$h(t) = \text{is-holiday}(t)$$

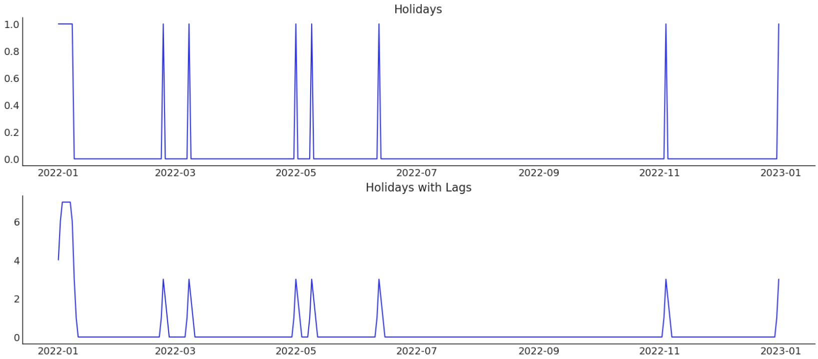


Figure: Признаки выходных дней

Периодическая часть (ковариация)

Здесь важна степень детализации. Можно использовать несколько периодических ядер.

- год
- квартал
- месяц
- неделя

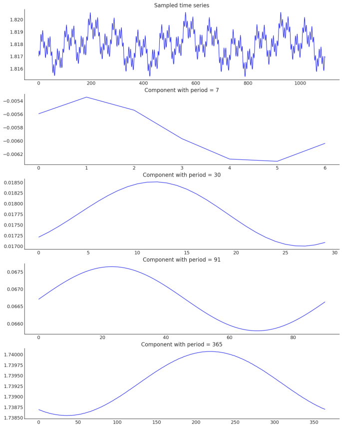


Figure: Сезонная декомпозиция

Масштабы длины для периодической части

Гиперпараметры

Выбор масштаба длины на основе здравого смысла

- Неделя – динамика может измениться за пару дней ($ls \approx 3$)
- Месяц – динамика может измениться за неделю ($ls \approx 7$)
- Квартал – динамика может измениться за месяц ($ls \approx 30$)
- Год – динамика может измениться за квартал ($ls \approx 90$)

Масштабы длины для периодической части

Гиперпараметры

Выбор масштаба длины на основе здравого смысла

- Неделя – динамика может измениться за пару дней ($ls \approx 3$)
- Месяц – динамика может измениться за неделю ($ls \approx 7$)
- Квартал – динамика может измениться за месяц ($ls \approx 30$)
- Год – динамика может измениться за квартал ($ls \approx 90$)

На практике

Вместо периодического ядра используют признаки Фурье (Fourier features).

Итоговая модель

$$m(t) = \underbrace{g(t)}_{\text{апериодичная часть}} + \underbrace{h(t)}_{\text{выходные}}$$

$$\begin{aligned} k(*,*) = {} & \mathsf{Periodic}(\mathsf{p} = 365, \mathsf{l} = 90) \\ + {} & \mathsf{Periodic}(\mathsf{p} = 90, \mathsf{l} = 30) \\ + {} & \mathsf{Periodic}(\mathsf{p} = 30, \mathsf{l} = 7) \\ + {} & \mathsf{Periodic}(\mathsf{p} = 7, \mathsf{l} = 3) \end{aligned}$$

Итоговая модель

$$m(t) = \underbrace{g(t)}_{\text{апериодичная часть}} + \underbrace{h(t)}_{\text{выходные}}$$

$$\begin{aligned} k(*,*) = {} & \alpha_{365}\,\mathrm{Periodic}(\mathrm{p} = 365, \mathrm{l} = 90) \\ & + \alpha_{90}\,\mathrm{Periodic}(\mathrm{p} = 90, \mathrm{l} = 30) \\ & + \alpha_{30}\,\mathrm{Periodic}(\mathrm{p} = 30, \mathrm{l} = 7) \\ & + \alpha_{7}\,\mathrm{Periodic}(\mathrm{p} = 7, \mathrm{l} = 3) \end{aligned}$$

Чего не хватает

1. Веса при периодических компонентах

Итоговая модель

$$m(t) = \underbrace{g(t)}_{\text{апериодичная часть}} + \underbrace{h(t)}_{\text{выходные}}$$

$$\begin{aligned} k(*,*) = {} & \alpha_{365}\,\mathrm{Periodic}(\mathrm{p} = 365, \mathrm{l} = 90) \\ & + \alpha_{90}\,\mathrm{Periodic}(\mathrm{p} = 90, \mathrm{l} = 30) \\ & + \alpha_{30}\,\mathrm{Periodic}(\mathrm{p} = 30, \mathrm{l} = 7) \\ & + \alpha_{7}\,\mathrm{Periodic}(\mathrm{p} = 7, \mathrm{l} = 3) \\ & + \beta\,\mathrm{ExpQuad}(\mathrm{l} = 90) + \mathrm{WhiteNoise}(\gamma) \end{aligned}$$

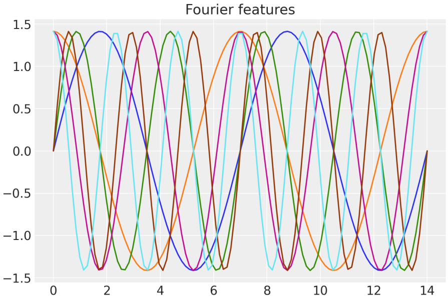
Чего не хватает

1. Веса при периодических компонентах
2. Отклонения от тренда $g(t)$

Более эффективная периодичность

Некоторые оптимизации недоступны, когда в модели есть хотя бы одно периодичное ядро.

- Можно добавить признаки Фурье как регрессоры
- Это позволяет добавить в модель разумную периодичность



Выбор признаков

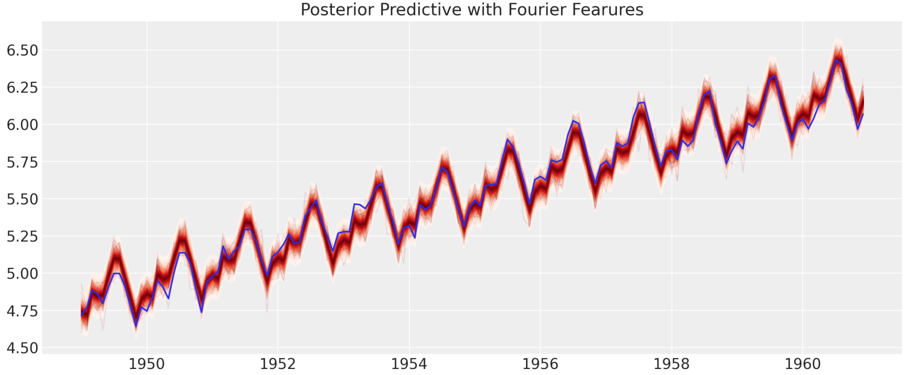
У каждого периода есть оптимальное число компонент Фурье (опрядок)

- неделя: 3
- месяц: 10
- год: 5

```
from collections import namedtuple
from enum import Enum

Season = namedtuple("Season", "period,order")

class Daily(Season, Enum):
    Week = 7.0, 3
    Year = 365.25, 5
    Month = 365.25 / 12, 10

class Monthly(Season, Enum):
    Year = 12, 5
    Month = 1, 10
```

Применение признаков Фурье

- Модели типа Prophet
- *Очень* быстрая и доступная альтертатива гауссовским процессам
- Полезны для байесовских моделей



Интеграция в модель

```
alpha = pm.Normal("alpha", 0, sigma, shape=Monthly.Year.order * 2)
features = fourier_series(months, season=Monthly.Year)
seasonality = features @ alpha

expectation = bias + trend + seasonality
```

Модель стохастической волатильности

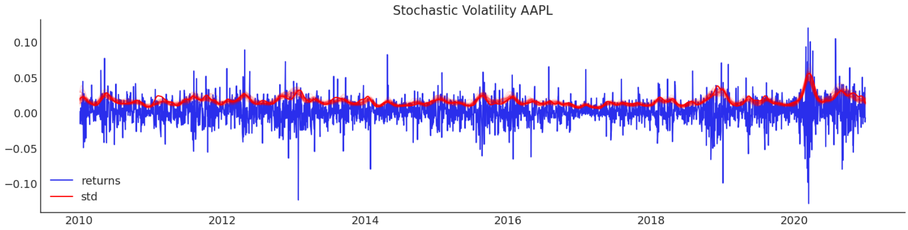


Figure: Оценка стохастической волатильности для AAPL

Подробнее

https://github.com/quantopian/bayesalpha

Основные идеи

- Доходности постоянны
- Волатильность не постоянна
- Моделируем волатильность как гауссовский процесс
- Используем приближения для ускорения модели
- Используем GPU для быстрого вывода

Априорные распределения

- Доходности

- Волатильность

- Временная компонента

Априорные распределения

- Доходности
  - Ожидаем годовую доходность порядка $\pm 100\%$
- Волатильность
- Временная компонента

Априорные распределения

- Доходности
  - Ожидаем годовую доходность порядка $\pm 100\%$
  - Дневная доходность $\pm(2^{\frac{1}{250}} - 1)$
- Волатильность
- Временная компонента

Априорные распределения

- Доходности
  - Ожидаем годовую доходность порядка $\pm 100\%$
  - Дневная доходность $\pm(2^{\frac{1}{250}} - 1)$
  - Априорное распределение безумно, но мотивировано, назовите своё!
- Волатильность
- Временная компонента

Априорные распределения

- Доходности
  - Ожидаем годовую доходность порядка $\pm 100\%$
  - Дневная доходность $\pm(2^{\frac{1}{250}} - 1)$
  - Априорное распределение безумно, но мотивировано, назовите своё!
- Волатильность
  - Ожидаем отношение сигнал/шум $\frac{\text{mean}}{\text{std}} \approx 1$
- Временная компонента

Априорные распределения

- Доходности
  - Ожидаем годовую доходность порядка $\pm 100\%$
  - Дневная доходность $\pm(2^{\frac{1}{250}} - 1)$
  - Априорное распределение безумно, но мотивировано, назовите своё!
- Волатильность
  - Ожидаем отношение сигнал/шум $\frac{\text{mean}}{\text{std}} \approx 1$
  - $\log \text{std} = \log(2^{\frac{1}{250}} - 1)$
- Временная компонента

Априорные распределения

- Доходности
  - Ожидаем годовую доходность порядка $\pm 100\%$
  - Дневная доходность $\pm(2^{\frac{1}{250}} - 1)$
  - Априорное распределение безумно, но мотивировано, назовите своё!
- Волатильность
  - Ожидаем отношение сигнал/шум $\frac{\text{mean}}{\text{std}} \approx 1$
  - $\log \text{std} = \log(2^{\frac{1}{250}} - 1)$
- Временная компонента
  - После построения модели

Модель

Я выбрал моделирование максимально простым

$$
\begin{aligned}
\text{return} &\sim \text{Normal}(0, 2^{\frac{1}{250}} - 1) \\
\log \text{std} &\sim \text{Normal}(\log(2^{\frac{1}{250}} - 1), 0.05) \\
\text{ls} &\sim \text{Gamma}(30, 5) \\
\alpha_{\text{vol}} &\sim \text{Exponential}(100) \\
K(*, *) &= \alpha_{\text{vol}} \, \text{Matern32}(\text{ls}) \\
\Delta_t^{\log \text{std}} &\sim \mathcal{GP}(0, K) \\
\text{obs}_t &\sim \text{Normal}(\text{return}, \exp(\log \text{std} + \Delta_t^{\log \text{std}}))
\end{aligned}
$$

Модель

Я выбрал моделирование максимально простым

$$
\begin{aligned}
\text{return} &\sim \text{Normal}(0, 2^{\frac{1}{250}} - 1) \\
\log \text{std} &\sim \text{Normal}(\log(2^{\frac{1}{250}} - 1), 0.05) \\
\text{ls} &\sim \text{Gamma}(30, 5) \\
\alpha_{\text{vol}} &\sim \text{Exponential}(100) \\
K(*, *) &= \alpha_{\text{vol}} \, \text{Matern32}(\text{ls}) \\
\Delta_t^{\log \text{std}} &\sim \mathcal{GP}(0, K) \\
\text{obs}_t &\sim \text{Normal}(\text{return}, \exp(\log \text{std} + \Delta_t^{\log \text{std}}))
\end{aligned}
$$

Модель

Я выбрал моделирование максимально простым

$$
\begin{aligned}
\text{return} &\sim \text{Normal}(0, 2^{\frac{1}{250}} - 1) \\
\log \text{std} &\sim \text{Normal}(\log(2^{\frac{1}{250}} - 1), 0.05) \\
\text{ls} &\sim \text{Gamma}(30, 5) \\
\alpha_{\text{vol}} &\sim \text{Exponential}(100) \\
K(*, *) &= \alpha_{\text{vol}} \,\text{Matern32}(\text{ls}) \\
\Delta_t^{\log \text{std}} &\sim \mathcal{GP}(0, K) \\
\text{obs}_t &\sim \text{Normal}(\text{return}, \exp(\log \text{std} + \Delta_t^{\log \text{std}}))
\end{aligned}
$$

Модель

Я выбрал моделирование максимально простым

$$
\begin{aligned}
\text{return} &\sim \text{Normal}(0, 2^{\frac{1}{250}} - 1) \\
\log \text{std} &\sim \text{Normal}(\log(2^{\frac{1}{250}} - 1), 0.05) \\
\text{ls} &\sim \text{Gamma}(30, 5) \\
\alpha_{\text{vol}} &\sim \text{Exponential}(100) \\
K(*, *) &= \alpha_{\text{vol}} \,\text{Matern32}(\text{ls}) \\
\Delta_t^{\log \text{std}} &\sim \mathcal{GP}(0, K) \\
\text{obs}_t &\sim \text{Normal}(\text{return}, \exp(\log \text{std} + \Delta_t^{\log \text{std}}))
\end{aligned}
$$

Модель

Я выбрал моделирование максимально простым

$$
\begin{aligned}
\text{return} &\sim \text{Normal}(0, 2^{\frac{1}{250}} - 1) \\
\log\text{std} &\sim \text{Normal}(\log(2^{\frac{1}{250}} - 1), 0.05) \\
\text{ls} &\sim \text{Gamma}(30, 5) \\
\alpha_{\text{vol}} &\sim \text{Exponential}(100) \\
K(*, *) &= \alpha_{\text{vol}}\,\text{Matern32}(\text{ls}) \\
\Delta_t^{\log\text{std}} &\sim \mathcal{GP}(0, K) \\
\text{obs}_t &\sim \text{Normal}(\text{return}, \exp(\log\text{std} + \Delta_t^{\log\text{std}}))
\end{aligned}
$$

Модель

Я выбрал моделирование максимально простым

$$
\begin{aligned}
\text{return} &\sim \text{Normal}(0, 2^{\frac{1}{250}} - 1) \\
\log \text{std} &\sim \text{Normal}(\log(2^{\frac{1}{250}} - 1), 0.05) \\
\text{ls} &\sim \text{Gamma}(30, 5) \\
\alpha_{\text{vol}} &\sim \text{Exponential}(100) \\
K(*, *) &= \alpha_{\text{vol}} \, \text{Matern32}(\text{ls}) \\
\Delta_t^{\log \text{std}} &\sim \mathcal{GP}(0, K) \\
\text{obs}_t &\sim \text{Normal}(\text{return}, \exp(\log \text{std} + \Delta_t^{\log \text{std}}))
\end{aligned}
$$

Модель

Я выбрал моделирование максимально простым

$$
\begin{aligned}
\text{return} &\sim \text{Normal}(0, 2^{\frac{1}{250}} - 1) \\
\log \text{std} &\sim \text{Normal}(\log(2^{\frac{1}{250}} - 1), 0.05) \\
\text{ls} &\sim \text{Gamma}(30, 5) \\
\alpha_{\text{vol}} &\sim \text{Exponential}(100) \\
K(*, *) &= \alpha_{\text{vol}} \, \text{Matern32}(\text{ls}) \\
\Delta_t^{\log \text{std}} &\sim \mathcal{GP}(0, K) \\
\text{obs}_t &\sim \text{Normal}(\text{return}, \exp(\log \text{std} + \Delta_t^{\log \text{std}}))
\end{aligned}
$$

Модель

Я выбрал моделирование максимально простым

$$
\begin{aligned}
\text{return} &\sim \text{Normal}(0, 2^{\frac{1}{250}} - 1) \\
\log \text{std} &\sim \text{Normal}(\log(2^{\frac{1}{250}} - 1), 0.05) \\
\text{ls} &\sim \text{Gamma}(30, 5) \\
\alpha_{\text{vol}} &\sim \text{Exponential}(100) \\
K(*, *) &= \alpha_{\text{vol}} \, \text{Matern32}(\text{ls}) \\
\Delta_t^{\log \text{std}} &\sim \mathcal{GP}(0, K) \\
\text{obs}_t &\sim \text{Normal}(\text{return}, \exp(\log \text{std} + \Delta_t^{\log \text{std}}))
\end{aligned}
$$

Ускорим модель!

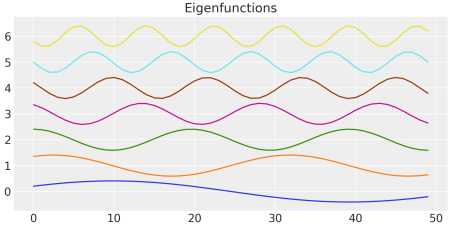
HSGP

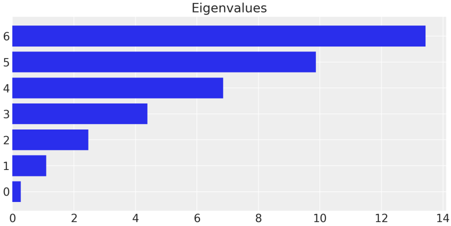
Что такое HSGP?

HSGP [1] – это сокращение для **H**ilbert **S**pace **G**aussian **P**rocess.

Часто задаваемые вопросы:

- Аппроксимирует в функциональном пространстве
- Работает для стационарных ядер
- Периодические ядра пока сложны
- Реализовано в PyMC





Зачем нужен HSGP?

1. Модели временных рядов имеют много временных точек
2. Сложность алгоритма растёт очень быстро ($\mathcal{O}(n^3)$)
3. HSGP [1] уменьшает сложность до $\mathcal{O}(mn + m)$
    - m - количество базисных функций)

Гауссовские процессы теперь быстрые!

Переход от $\mathcal{O}(n^3)$ к $\mathcal{O}(mn + m)$ имеет большой смысл!

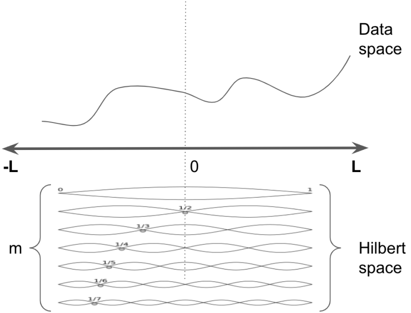
Как использовать HSGP

Есть два основных параметра:

- L - граница экстраполяции
- m - количество базисных функций

Вы можете использовать это

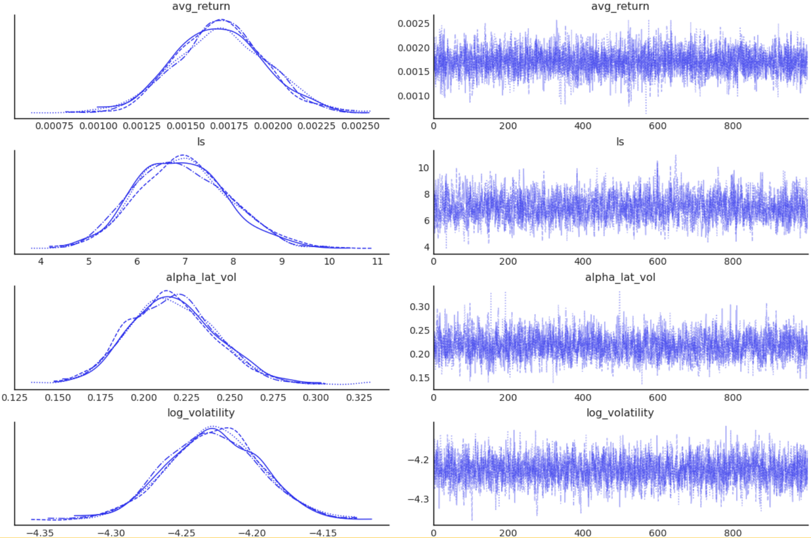
Не нужно быть математиком, чтобы задать L или m, в документации есть хорошие значения по умолчанию.



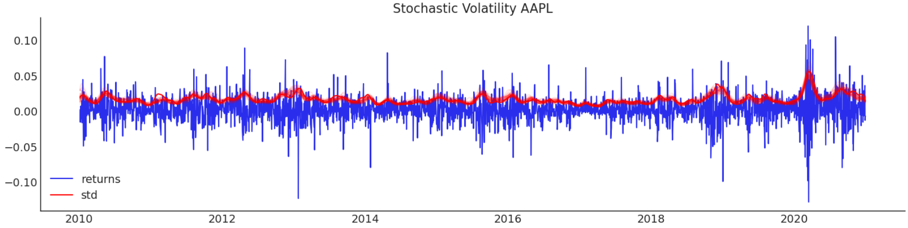
HSGP для стохастической волатильности

$$
\begin{aligned}
\text{return} &\sim \text{Normal}(0, 2^{\frac{1}{250}} - 1) \\
\log \text{std} &\sim \text{Normal}(2^{\frac{1}{250}} - 1, 0.05) \\
\text{ls} &\sim \text{Gamma}(30, 5) \\
\alpha_{\text{vol}} &\sim \text{Exponential}(100) \\
K(*, *) &= \alpha_{\text{vol}} \, \text{Matern32}(\text{ls}) \\
\Delta_t^{\log \text{std}} &\sim \mathcal{HSGP}(0, K) \\
\text{obs}_t &\sim \text{Normal}(\text{return}, \exp(\log \text{std} + \Delta_t^{\log \text{std}}))
\end{aligned}
$$

Результаты


avg_return
0.00075 0.00100 0.00125 0.00150 0.00175 0.00200 0.00225 0.00250
avg_return
0.0025
0.0020
0.0015
0.0010
0 200 400 600 800
ls
4 5 6 7 8 9 10 11
ls
10
8
6
4
0 200 400 600 800
alpha_lat_vol
0.125 0.150 0.175 0.200 0.225 0.250 0.275 0.300 0.325
alpha_lat_vol
0.30
0.25
0.20
0.15
0 200 400 600 800
log_volatility
−4.35 −4.30 −4.25 −4.20 −4.15
log_volatility
−4.2
−4.3
0 200 400 600 800

Выводы



- Интерпретируемые параметры
- Сплайны делают модель намного быстрее
- Неопределённости для
  - средней волатильности
  - средних доходностей
  - стохастической вариации (α_{vol})

Библиография I

G. Riutort-Mayol, P.-C. Bürkner, M. R. Andersen, A. Solin, and A. Vehtari.
Practical hilbert space approximate bayesian gaussian processes for probabilistic programming, 2022.

T. SJ and L. B.
Forecasting at scale.
2017.