Регрессия на основе Гауссовских процессов

Алексей Зайцев
Руководитель
лаборатории
Сколтех

Skoltech

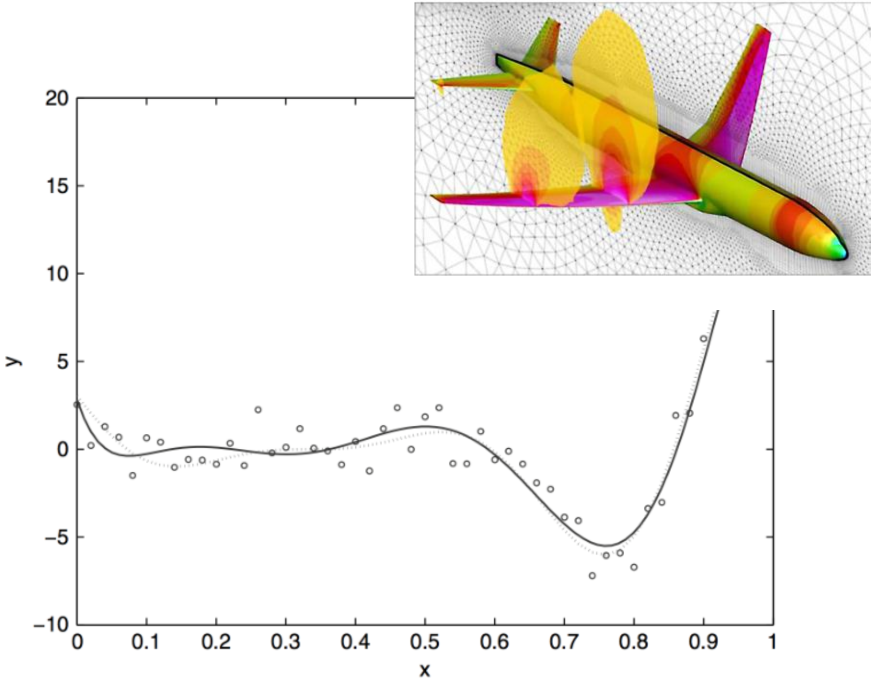
Про что лекция

1. Мотивация - что нам нужно от суррогатной модели
2. Ядерная регрессия, kNN регрессия
3. Что такое многомерное гауссовское распределение, условное гауссовское распределение
4. Примеры случайных процессов, что такое случайный процесс, что такое гауссовский случайный процесс
5. Регрессия на основе гауссовских процессов, функция среднего и ковариационная функция
6. Формулы для оценки среднего и МО, оценка параметров методом максимума правдоподобия
7. Гауссовские процессы для выборок большого размера, глубокие гауссовские процессы

Skoltech

Мотивация и постановка задачи суррогатного моделирования

Нужно оптимизировать тяжелую функцию в многомерном пространстве, поэтому можно сделать ограниченное количество в запусков.



Функция обычно:

- Тяжелая
- Гладкая
- Но наблюдаем мы ее с шумом

Примеры:

- Оптимизация гиперпараметров модели

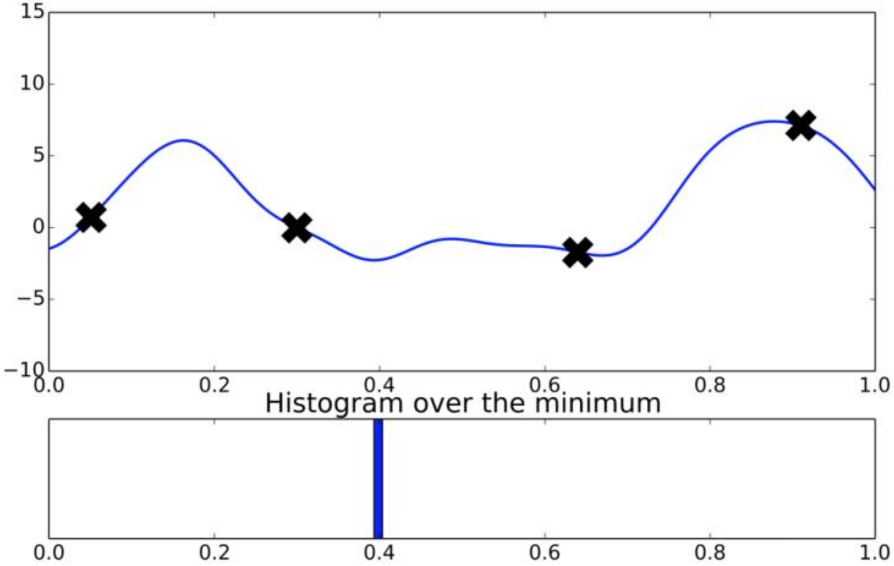
Как максимизировать функцию с помощью суррогатной модели

1. Генерируем начальную выборку

2. Учим суррогатную регрессионную модель на этой выборке

3. Оптимизируем суррогатную модель вместо исходной тяжелой модели

Проблема: очень жадный (локально оптимальный) подход



Как такую оптимизацию устроить

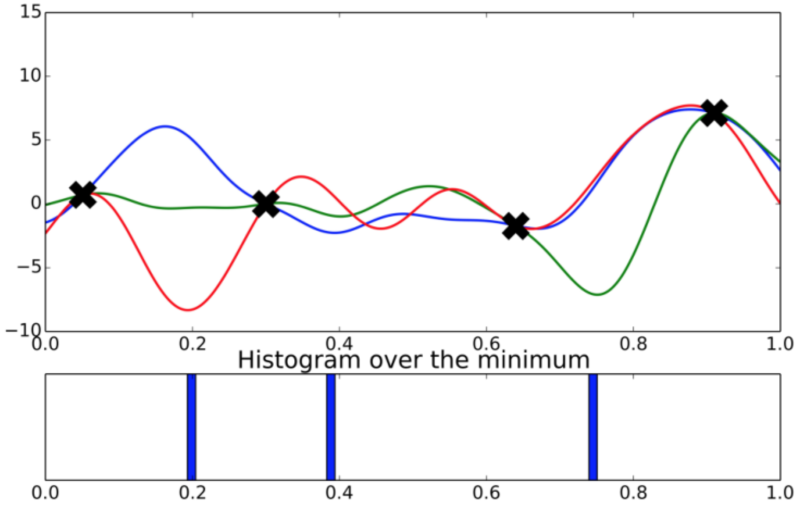
1. Генерируем начальную выборку

2. Учим суррогатную регрессионную модель на этой выборке

3. Оптимизируем суррогатную модель вместо исходной тяжелой модели, получили еще одну точку

4. Переучили суррогатную модель

Проблема: все равно жадный подход



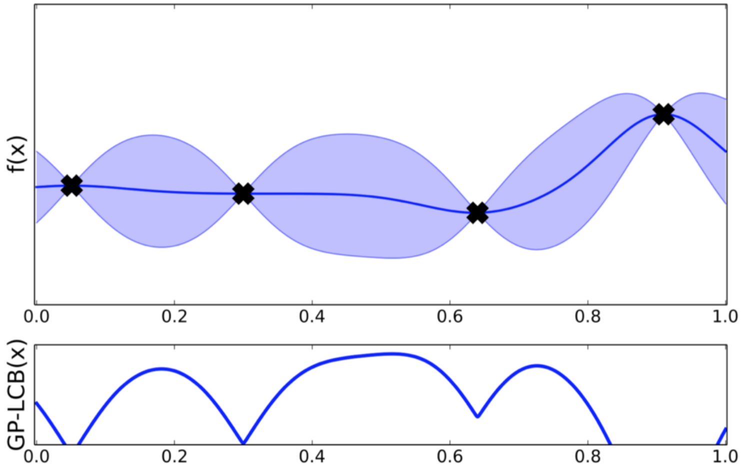
Skoltech

Алгоритм: как такую оптимизацию устроить

1. Генерируем начальную выборку
2. Учим суррогатную регрессионную модель на этой выборке
3. Оптимизируем суррогатную модель + ее неопределенность вместо исходной тяжелой модели, получили еще одну точку
4. Переучили суррогатную модель
5. Повторяем пока не закончится бюджет

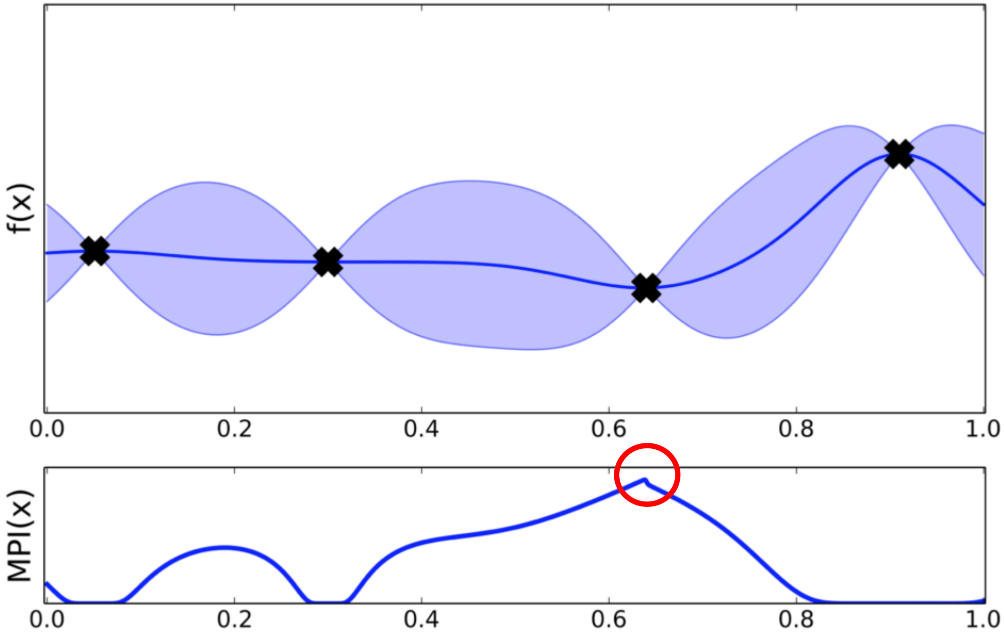
Проблемы нет: теперь работает нормально

$$\alpha_{LCB}(\mathbf{x}) = -\mu_*(\mathbf{x}) + \zeta \cdot \sigma_*(\mathbf{x})$$



Еще пример, как не работает

Вероятность улучшения в качестве целевой функции



Skoltech

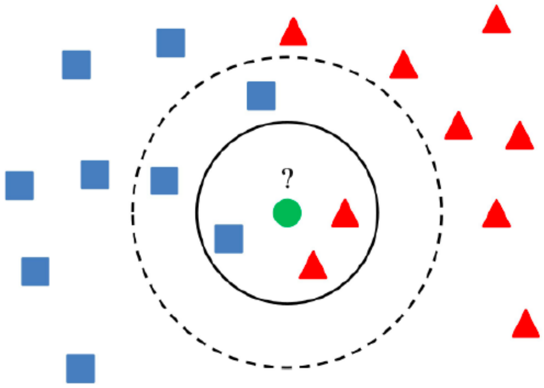
kNN регрессия и ядерная регрессия

kNN регрессия: усредняем по ближайшим соседям

Ядерная регрессия: усредняем с весами, веса пропорциональны расстоянию до объекта

Чего нам не хватает:

- Нет учета знаний
- Как выбрать ядро и число соседей?
- Как оценить неопределенность модели?

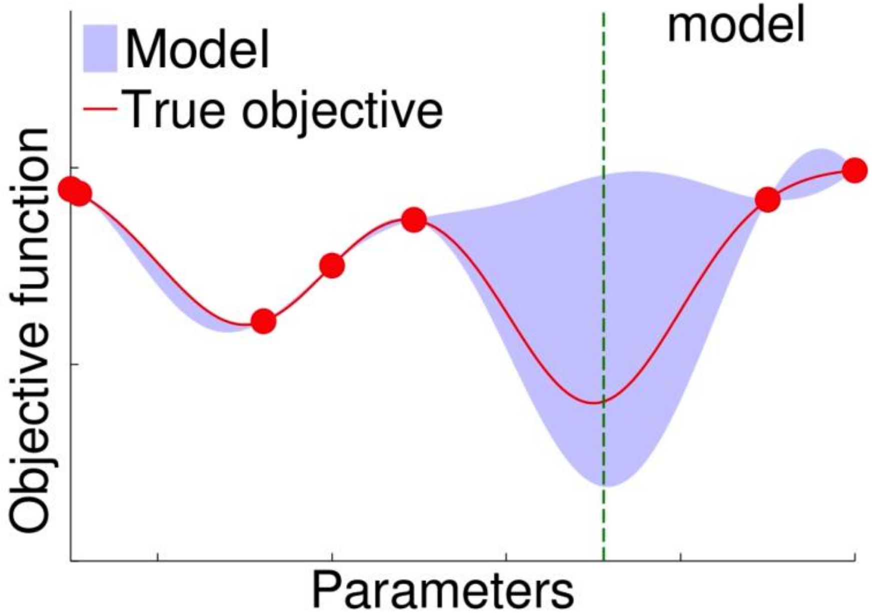


А что нам нужно?

- Нелинейность, локальность, интерполяция
- Оценка неопределенности
- Гладкость
- Быстро переучивать
- Хорошо работает для маленьких выборок

Даст ли нам это:

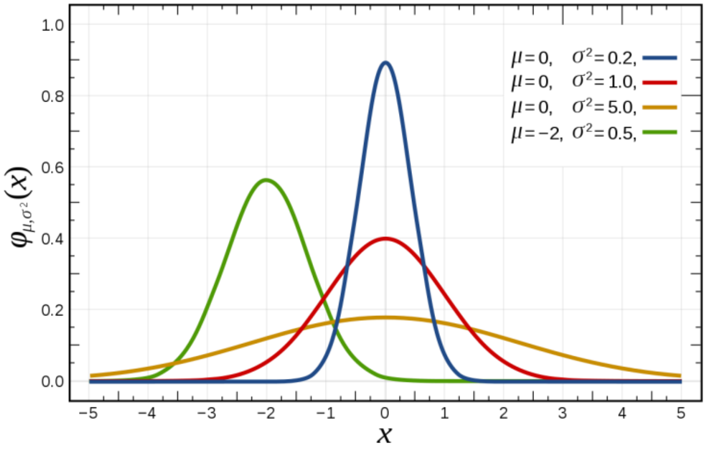
- Линейная регрессия
- Решающий лес
- **Регрессия на основе гауссовских процессов**



Skoltech

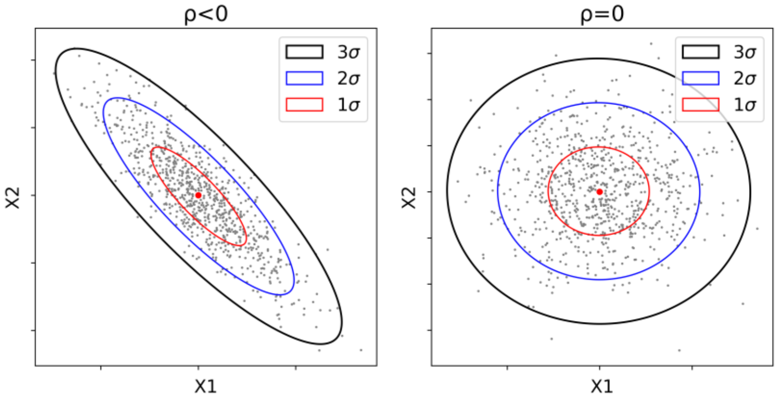
Многомерное гауссовское распределение

Плотность одномерного гауссовского распределения



$$p(x;\mu,\sigma^2)=\frac{1}{\sqrt{2\pi}\,\sigma}\exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$$

Плотность многомерного гауссовского распределения

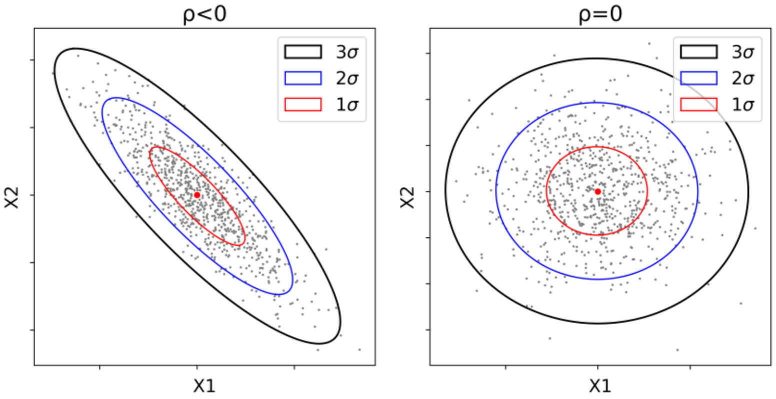


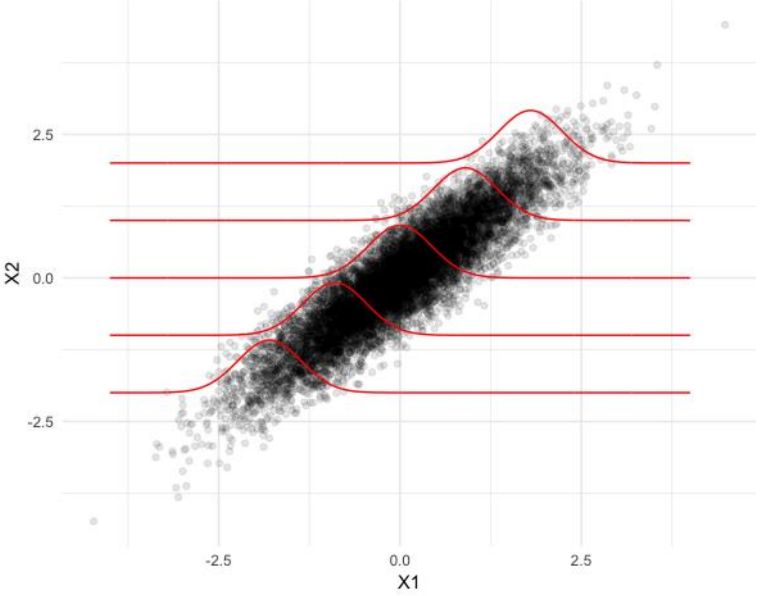
$$p(\boldsymbol{x};\boldsymbol{\mu},\Sigma)=\frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

Skoltech

Условное гауссовское распределение

Плотность многомерного
гауссовского распределения





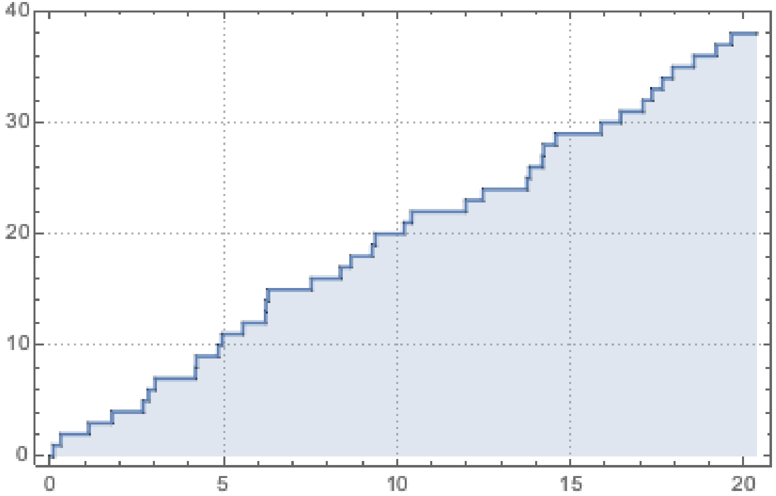
$$p(\boldsymbol{x};\boldsymbol{\mu},\Sigma)=\frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{T}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right)$$

Skoltech

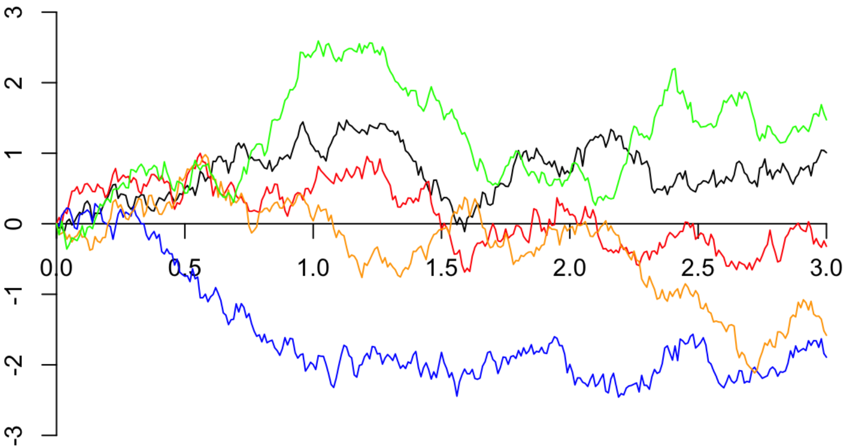
Примеры случайных процессов

Примеры случайных процессов

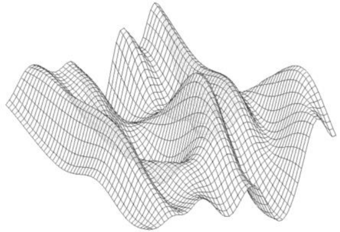
Пуассоновский процесс



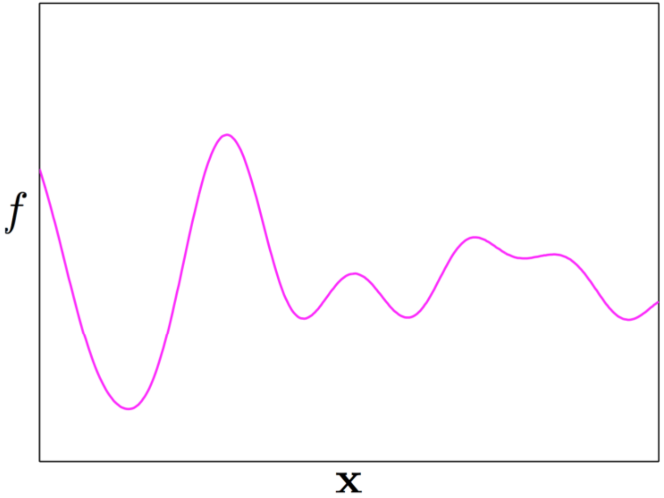
Броуновское движение

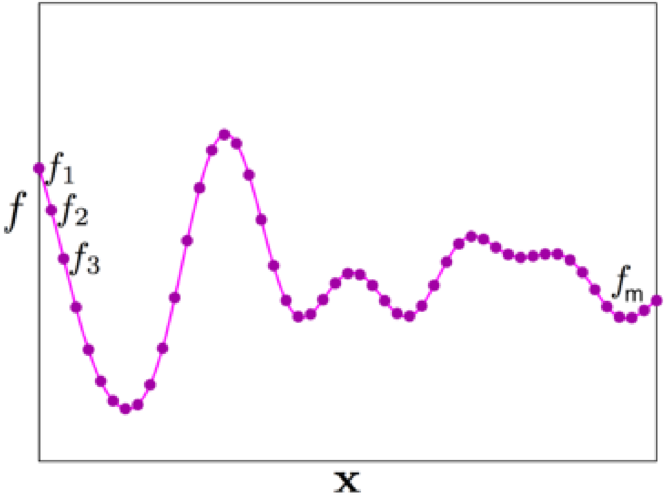


Гауссовский случайный процесс



Совместное распределение для любого множества точек - нормальное



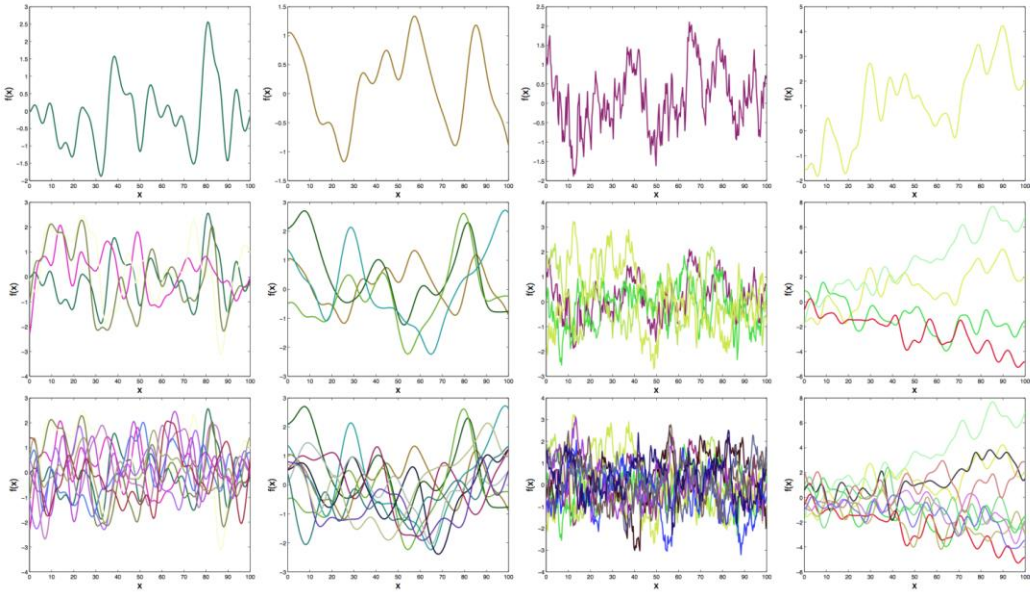


$$f(\mathbf{x}) \sim \mathcal{GP}(\cdot|\mu(\mathbf{x}), K(\mathbf{x}, \mathbf{x}'))$$

Skoltech

Регрессия на основе гауссовских процессов

$$K(\mathbf{x}, \mathbf{x}') = \sigma_f^2 \exp\left\{-\sum_{i=1}^{d} \frac{(x_i - x_i')^2}{2r_i^2}\right\}$$



Оценка среднего и дисперсии для регрессии на основе гауссовских процессов

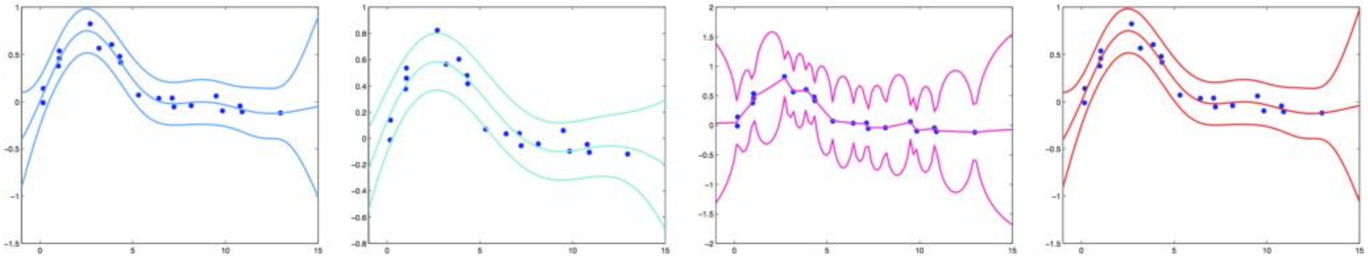
Аналитические формулы для среднего и дисперсии:

$$p(f_*|\mathbf{y}) = \mathcal{N}\left(f_*|\mu_*, \sigma_*^2\right),$$

$$\mu_* = \mathbf{k}_*^{\mathrm{T}}[\mathbf{K} + \sigma^2 \mathrm{I}_m]^{-1}\mathbf{y}, \;\; \sigma_*^2 = K_{**} - \mathbf{k}_*^{\mathrm{T}}[\mathbf{K} + \sigma^2 \mathrm{I}_m]^{-1}\mathbf{k}_*$$

$$\mathbf{k}_* = \{K(\mathbf{x}_*, \mathbf{x}_i)\}_{i=1}^m \text{ and } K_{**} = K(\mathbf{x}_*, \mathbf{x}_*)$$

$$\mu_* = \sum_{i=1}^m \alpha_i K(\mathbf{x}_*, \mathbf{x}_i), \;\; \boldsymbol{\alpha} = [\mathbf{K} + \sigma^2 \mathrm{I}_m]^{-1}\mathbf{y}$$



Skoltech

Метод максимума правдоподобия для оценки параметров

$$\mathcal{L} = -\log p(\mathbf{y}|\boldsymbol{\theta}) = \underbrace{\frac{1}{2}\log\det\mathbf{C}(\boldsymbol{\theta})}_{\color{red}{regularization}} + \underbrace{\frac{1}{2}\mathbf{y}^{\mathrm{T}}\mathbf{C}^{-1}(\boldsymbol{\theta})\mathbf{y}}_{\color{red}{data-fit}} + \frac{m}{2}\log(2\pi),$$

- Умеем считать правдоподобие
- Умеем считать производные
- Запускаем градиентный спуск до сходимости, часто глобальный оптимум один

Skoltech

Проблемы гауссовских процессов

A. Большая вычислительная сложность, для размера выборки m:

- Оценка правдоподобия $O(m^3)$
- Прогнозирование $O(m)$
- Оценка неопределенности $O(m^2)$

B. Проклятье размерности: Большая размерность приводит к плохим результатам для ядерных методов

Решение 1: РГП для выборок большого размера

- Приближенные методы для обращения ковариационной матрицы

- Метод на основе признаков Фурье, так как РГП эквивалентно линейной регрессии в бесконечномерном пространстве

$$K(\mathbf{x}, \mathbf{x}') \approx \frac{1}{k} \sum_{i=1}^{k} \phi(\mathbf{x}, \mathbf{w}_i)\phi(\mathbf{x}', \mathbf{w}_i) = \langle \psi(\mathbf{x}), \psi(\mathbf{x}') \rangle, \;\; \mathbf{w}_i \sim p(\mathbf{w})$$

- Вместо O(m^3) - O(m)!

Решение 2: глубокая регрессия на основе гауссовских процессов

- Построить регрессию на основе гауссовских процессов на представлениях или для проекций
- *У нас работало для: нейросетевых архитектур, молекул*

- Сделать полный Байесовский вывод
- *С точки зрения теории хороший подход, на практике не сильно лучше первого варианта*

Зачем нам это нужно:
- Оценка неопределенности
- Более точная модель

Skoltech

Выводы

Выводы

- Выбор класса моделей для построения суррогатной модели – непростая задача
- Регрессия на основе гауссовских процессов решает поставленную задачу
- Однако, у нее есть недостатки – она не работает с неструктурированными данными, большими размерностями и выборками
- Современные подходя позволяют с этими недостатками бороться