

Московский физико-технический институт
(национальный исследовательский университет)



Название потом придумаю

Выполнил:
студент 1 курса ЛФИ
Сливка Глеб

Содержание

1	Теоретические сведения	2
1.1	Необходимы определения машинного обучения	2
1.2	Гауссовский процесс	3
1.3	Задача нелинейной регрессии	3
1.4	Оптимизация	4
2	Постановка задачи	5

1 Теоретические сведения

1.1 Необходимы определения машинного обучения

Опр. *Гиперпараметр* — параметр машинного обучения, значение которого используется для управления процессом обучения. Его значение устанавливается перед началом обучения, в отличие от значений других параметров (обычно весов узлов), которые определяются во время обучения.

Опр. *Операция оценки функции* — это процесс определения, насколько хорошо модель машинного обучения способна предсказывать значения целевой переменной на основе входных данных.

Оценка функции производится путём сравнения предсказанных значений модели с фактическими значениями целевой переменной. Чем ближе предсказания модели к фактическим значениям, тем выше оценка функции и тем лучше модель способна выполнять предсказания.

Опр. *Функция стоимости* — это функция, которая измеряет стоимость или ошибку модели в зависимости от ее параметров. Она используется для оптимизации модели путем настройки ее параметров таким образом, чтобы минимизировать стоимость или ошибку.

Опр. *Функция потерь* — это функция, которая измеряет разницу между предсказанными значениями модели и фактическими значениями целевой переменной. Она используется для оценки качества модели и настройки ее параметров.

Функция потерь определяет, насколько хорошо модель выполняет задачу, которую она должна решать. Чем меньше значение функции потерь, тем лучше модель соответствует данным и делает более точные предсказания.

Опр. *Байесовская оптимизация* — это метод оптимизации функций, который использует байесовский подход для поиска оптимальных значений гиперпараметров модели.

В машинном обучении, модель может иметь гиперпараметры, которые не могут быть обучены напрямую из данных, и требуется выбрать оптимальные значения для этих гиперпараметров. Байесовская оптимизация позволяет находить оптимальные значения гиперпараметров, минимизируя количество итераций обучения модели.

То есть байесовская оптимизация находит такое $x' \in \mathbb{R}^n$ — набор неизвестных гиперпараметров на данном множестве X , что функция f принимает максимальное значение:

$$x' = \underset{x \in X}{\operatorname{argmax}} [f(x)].$$

На каждой итерации алгоритм может вычислить оптимальный x и сверить его с реальным тем самым можно *производить операцию оценки функции*.

1.2 Гауссовский процесс

Рассмотри индекс $x' \in \mathbb{R}^n$ и гауссовский процесс $f(x)$. Определим функцию среднего

$$m(x) = \mathbb{M}f(x),$$

которая показывает среднее значение функции, и функцию ковариации

$$k(x, x') = \mathbb{M}[(f(x) - m(x))(f(x') - m(x'))],$$

которая показывает как изменение одной переменной связано с изменением другой переменной.

Опр. $f(x)$ — *Гауссовский процесс (ГП)*, если $\forall n \in \mathbb{N}$ и $\forall x_1, x_2, \dots, x_n$ величина $X = (f(x_1), \dots, f(x_n))$ имеет многомерное гауссово распределение с вектором среднего $\mathbb{M}x = (m(x_i))_{i=1}^n$ и матрицей ковариации $\mathbb{D}x = (k(x_i, x_j))_{i,j}$.

Мы будем работать со стационарными ГП, поэтому $m(x) = \text{const}$. Если положить $m(x) = 0$, то ничего в задаче не поменяется.

Самый распространённый выбор функции корреляции (ядра ГП) — экспоненциальное ядро:

$$k(x, x') = \alpha^2 \cdot \exp \left[-\frac{1}{2\lambda} (x - x')^T (x - x') \right],$$

где α, λ — гиперпараметры ядра.

1.3 Задача нелинейной регрессии

Пусть мы знаем значение функции f в точках x_1, x_2, \dots, x_n и $(f(x_1), f(x_2), \dots, f(x_n)) = (f_1, f_2, \dots, f_n) = f$. Задача нелинейной регрессии сводится к поиску значению функции f в произвольной фиксированной точке x' .

Так как $f(x')$ — случайная величина, то плотность вероятности по формуле Байеса вычисляется следующим образом:

$$p(f(x')|f_1, \dots, f_n) = \frac{p(f(x'), f_1, \dots, f_n)}{p(f_1, \dots, f_n)} = \mathcal{N}(f(x)|\mu, \sigma^2).$$

Мы получили нормальное распределение. Определим матрицу K :

$$K = \begin{pmatrix} C & k \\ k^T & k(0) \end{pmatrix},$$

где

$$C = (k(x_i, x_j))_{i,j=1}^n, \quad k_i = k(x_i, x').$$

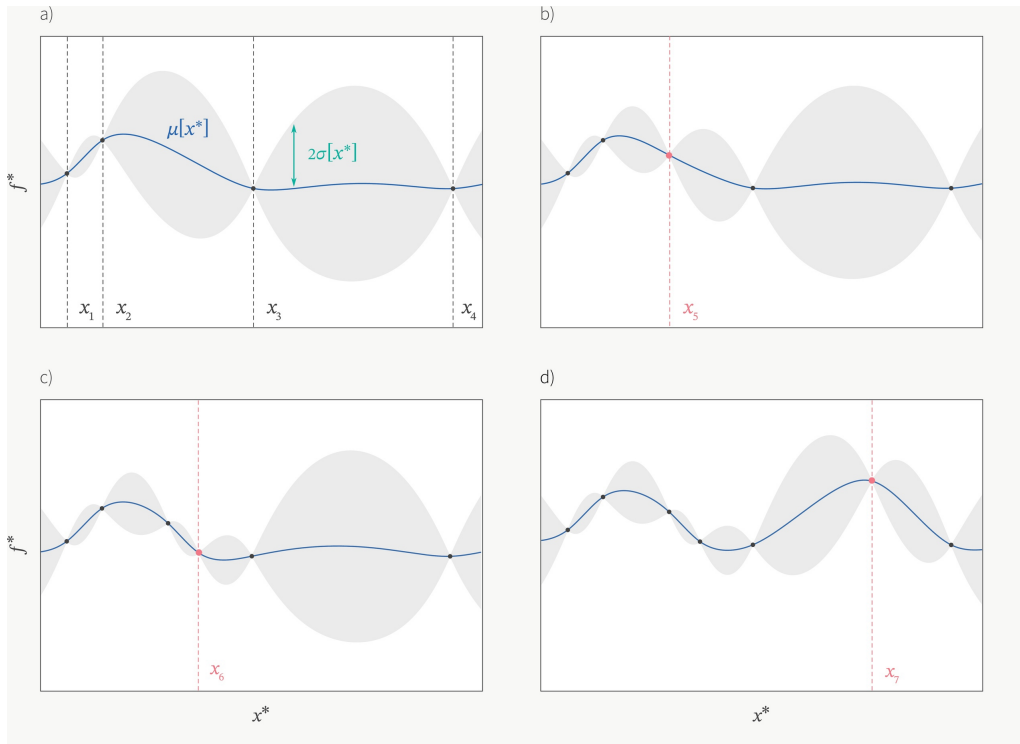
Тогда математическое ожидание и дисперсия записываются в следующем виде:

$$\mu = k^T C^{-1} f,$$

$$\sigma^2 = k(0) - k^T C^{-1} k.$$

Используя эти формулы можно найти распределение функции в каждой новой точке.

Пусть у нас есть выборка из четырёх точек, тогда ν будет проходить через все эти точки, и σ^2 будет определять доверительный диапазон. На рисунке ниже показано, как будет менять функция при добавление новых точек в выборку.



Также стоит отметить, что если x' — точка далёкая от точек выборки, то есть $k(x, x') \rightarrow 0$, тогда вектор $k \rightarrow 0$, значит функция стремится к 0. То есть, если мы обладаем конечной выборкой, то на бесконечности функция будет стремиться к 0.

Так как на бесконечности $k \rightarrow 0$, то $\sigma^2 \rightarrow k(0)$, то есть доверительный интервал будет конечным и постоянным.

1.4 Оптимизация

Пусть у нас есть выборка $y(x_1), \dots, y(x_n)$ для оптимизируем функции. Решим задачу регрессии с помощью гауссовского процесса и получим некоторую функцию $f(x)$. Чтобы оценить, насколько полезно будет исследовать конкретную точку в пространстве параметров вводят Acquisition функцию.

Она основывается на текущей модели целевой функции и может учитывать различные факторы, такие как неопределённость модели (например, доверительный интервал), предполагаемое улучшение (например, насколько сильно значение целевой функции может быть улучшено в данной точке) и баланс между исследованием и использованием уже известной информации.

Acquistion функции принимают среднее и дисперсию в каждой точке x функции и вычисляют значение, которое описывает насколько желательно выбрать эту точку в следующий раз.

Модно придумать много разных Acquistion функций, мы рассмотрим следующую функцию:

$$f_{AC}(\theta) = \alpha\mu(\theta) + \beta\sigma(\theta),$$

где $\mu(\theta)$ — значение, предсказанное ГП в точке θ со стандартным отклонением $\sigma(\theta)$, α и β — фиксированные во время обучения коэффициенты.

Обобщим алгоритм оптимизации функции:

1. Есть выборка точек x_1, \dots, x_n , значения оптимизируемой функции в которых мы знаем.
2. «Фитим» функцию по данным точка с помощью ГП.
3. Вычисляем значение целевой функции в точке $x' = \underset{x' \in X}{argmax}[f_{AC}(x)]$.
4. Добавляем $x_{n+1} = x'$ (gaussian process) или случайную точку (rand), чтобы не застревать в локальных минимумах функции.

2 Постановка задачи

В нашем случае нужно оптимизировать параметры испарительного охлаждения в ОДЛ. В качестве гиперпараметров, которые оптимизировали с помощью байесовской оптимизации взяли мощности горизонтального и вертикального пучков и ширину горизонтального пучка. Изменение мощности и ширины пучков во времени задаётся линейной аппроксимацией. После прохождения цикла охлаждения с заданными параметрами делается фотография, из которой можно достать всю необходимую информацию (количество атомов, температуру и тд). По этим данным высчитывается функция стоимости, минимум которой необходимо найти.