

Суррогатная модель (Surrogate model)

Синонимы: Мета модель, Эмулятор, Подменная модель, Замещающая модель, Аппроксимирующая модель, Metamodel, Emulator

Суррогатными называют упрощенные и недорогие модели, которые используются для замещения более сложных, точных и дорогих имитационных, затраты на построение и эксплуатацию которых неприемлемы при решении конкретной задачи. Изначально суррогатное моделирование использовалось главным образом в инженерном проектировании, однако постепенно проникло и в другие сферы деятельности, где требуется применение моделей.

Суррогатная модель описывает не сам объект, а аппроксимирует некоторую более сложную и точную модель этого объекта. При этом целевая функция суррогатной впоследствии замещает целевую функцию точной модели. Иными словами, суррогатное моделирование подразумевает замену функции $f(x)$, которая является слишком сложной и дорогой для оценки, на более простую $g(x)$, имеющую такое же количество входных и выходных переменных, что и исходная функция. Суррогатная модель должна аппроксимировать поведение точной модели максимально близко, насколько это возможно при приемлемых затратах.



При моделировании сложных систем, какими являются технические, социально-экономические и бизнес системы, может потребоваться оценка большого числа параметров и проведение множества экспериментов. И зачастую проектировщики и

аналитики сталкиваются с ситуацией, когда издержки использования точной модели превышают выгоду.

Например, стоит задача спрогнозировать спрос на новый продукт компании, чтобы оценить объемы поставок и цену. Если процесс построения прогнозной модели окажется слишком долгим, то конкуренты успеют вывести на рынок свои аналогичные продукты и смогут получить преимущество без всякого моделирования.

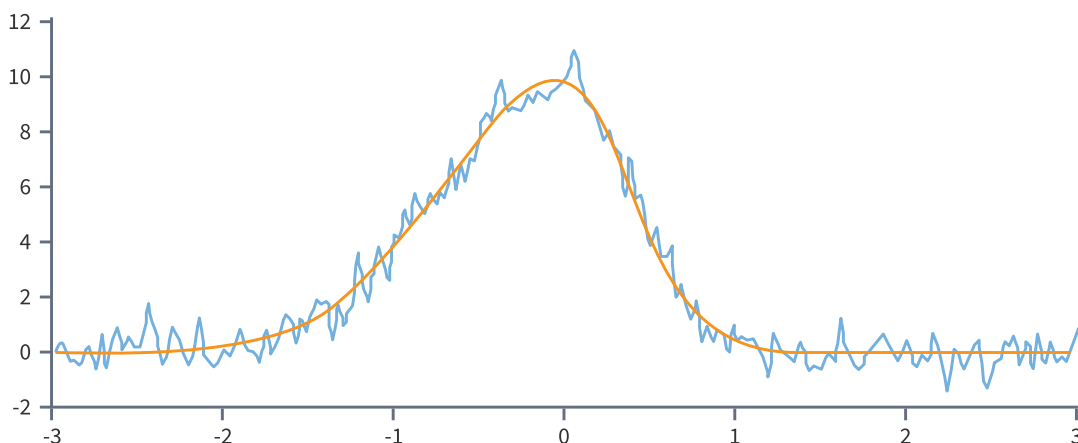
Именно в этой ситуации и призвано помочь суррогатное моделирование — замещение точной, сложной и дорогой модели на более простую и дешевую, хотя и менее точную, но позволяющую получить результат за приемлемое время и с разумными затратами. Иными словами, суррогатная модель должна обеспечивать максимальную точность при как можно меньшем числе оценок параметров.

Как правило, в процессе моделирования требуется решить следующие задачи:

- **анализ чувствительности** — спрогнозировать поведение объекта при изменении его характеристик или параметров внешнего окружения (например изменение спроса на новый продукт при изменении его цены и курса валют);
- **оптимизация параметров** — выбор из множества возможных конфигураций объекта той, которая обеспечит наилучшие свойства (например, для продукта минимальную цену, наибольшую прибыль и т.д.);
- **анализ рисков** — оценить риски того, что объект не оправдает наши ожидания (скажем, продукт провалится на рынке, принесет убытки и т.д.).

Решение этих задач требует большого числа итераций моделирования, на каждой из которых в качестве входных данных будут использоваться различные комбинации атрибутов самого продукта и внешнего окружения. При этом внутренняя структура модели и механизмы ее работы не существенны, важно только изменение выхода при изменении входа. Т.е. суррогатная модель строится по принципу черного ящика.

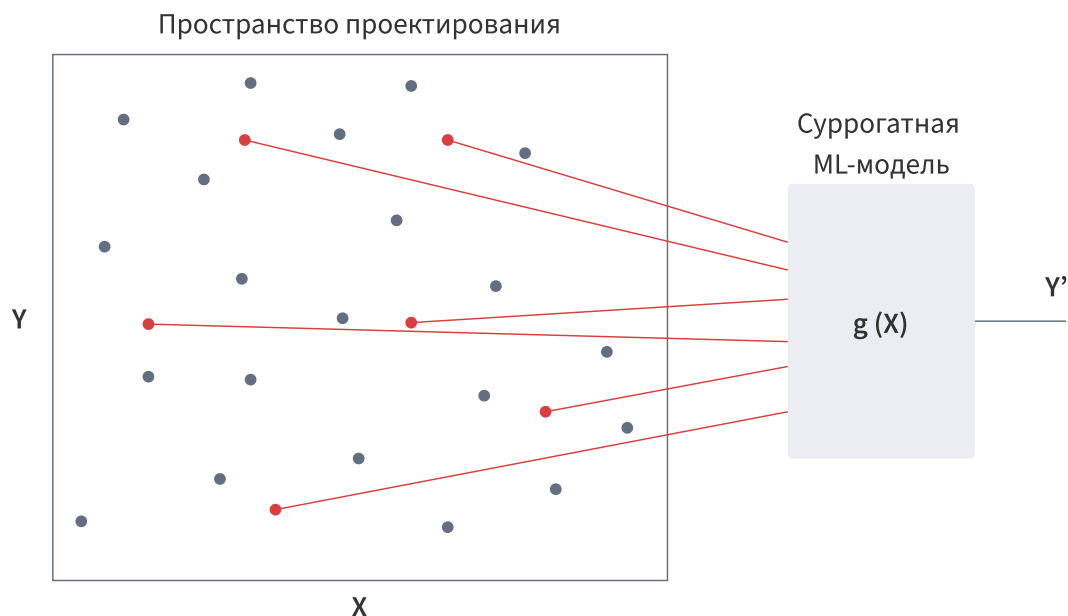
Традиционно технологии суррогатного моделирования использовали статистические методы: метод подгонки кривой, метод поверхности отклика, кригинг, пространственное отображение, байесовский подход. Например, метод подгонки кривой можно проиллюстрировать с помощью рисунка.



В этом случае сложная функция с шумом аппроксимируется более простой и гладкой.

Однако в последнее время популярность приобретают методы суррогатного моделирования, использующие машинное обучение, такие как нейронные сети, машины опорных векторов, случайные леса и т.д.

Данные для обучения получают путем анализа точной модели и ее выполнения в ограниченном числе точек пространства проектирования. При этом каждая комбинация пар вход-выход точной модели используется в качестве обучающего примера для суррогатной.



На рисунке схематически показано пространство моделирования, каждая точка которого представляет собой некоторую оценку параметров моделируемого объекта или процесса. Для получения точной модели нужно произвести оценивание во всех точках. Если это очень долго и дорого, то оценка производится на каком-то ограниченном числе точек (обозначены красным цветом). Из них и формируется обучающий набор данных.

Построение суррогатных моделей на основе машинного обучения содержит следующие шаги:

- 1. Формирование обучающего набора данных.** В пространстве проектирования выбирается определенное количество областей, соответствующих интервалам значений параметров объекта (например, доходы клиентов от 50 до 70 тыс. рублей и возраст от 30 до 40 лет). Затем из каждой области выбирается одна или несколько точек для оценивания. Выбор областей производится на основе опросов экспертов, с использованием методов планирования эксперимента или просто равномерно.
- 2. Определение целевых значений для обучения.** Производится прогон точной модели на выбранных точках оценивания, и результаты используются как целевые значения для алгоритмов обучения с учителем.
- 3. Выбор вида, архитектуры и конфигурации ML-модели и обучение на сформированном наборе данных.** Как правило, для этого требуется множество

экспериментов. Рекомендуется начать с наиболее простых моделей и их конфигураций, постепенно усложняя, до применения ансамблей, чтобы построить модель максимально близкую к точной.

4. **Активное обучение.** Поскольку число обучающих примеров, позволяющих построить хорошую ML-модель, заранее неизвестно, то начинают с относительно небольшого набора и постепенно увеличивают их количество до получения приемлемого результата. Данный подход называется **активным обучением**. При этом новые точки-кандидаты выбирают так, чтобы они были наиболее информативными.
5. **Обогащение обучающей выборки.** Точная модель прогоняется на новых точках, формируя целевые значения, и обучающая выборка обогащается новыми примерами.

Пункты 4 и 5 итеративно повторяются до получения приемлемой точности суррогатной модели.

Таким образом, суррогатное моделирование делает дорогостоящий анализ данных более доступным. Применение суррогатных моделей позволяет сократить стоимость и время моделирования на 30-50%, что обеспечивает значительные конкурентные преимущества.