МОСКОВСКИЙ АВИАЦИОННЫЙ ИНСТИТУТ

(НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ)

Институт №8 «Информационные технологии и прикладная математика» Кафедра 804 «Теория вероятностей и компьютерное моделирование»

Реферат по курсу

“Эконометрика”

По теме
“Непараметрическая регрессия”

Выполнил: Дюсекеев А.Е.

Группа: 8О-404Б

Москва, 2020

Непараметрическая регрессия - это категория регрессионного анализа, в которой предиктор не принимает заранее заданную форму, а строится в соответствии с информацией, полученной из данных. То есть не предполагается параметрической формы отношения между предикторами и зависимой переменной. Непараметрическая регрессия требует больших размеров выборки, чем регрессия, основанная на параметрических моделях, потому что данные должны предоставлять структуру модели, а также оценки модели.

Непараметрическая регрессия отличается от параметрической регрессии тем, что форма функциональных отношений между ответной (зависимой) и объясняющей (независимой) переменными не предопределена заранее, но может быть скорректирована для выявления необычных или неожиданных характеристик данных. Когда взаимосвязь между ответом и независимыми переменными известна, следует использовать модели параметрической регрессии. Если взаимосвязь неизвестна и нелинейна, следует использовать непараметрические модели регрессии. В случае, если мы знаем взаимосвязь между ответом и частью объясняющих переменных и не знаем взаимосвязи между ответом и другой частью объясняющих переменных, мы используем модели полупараметрической регрессии. Любая прикладная область, в которой используется регрессионный анализ, потенциально может выиграть от полу / непараметрической регрессии.

Объем непараметрической регрессии очень широк: от «сглаживания» взаимосвязи между двумя переменными на диаграмме рассеяния до анализа множественной регрессии и обобщенных регрессионных моделей (например, логистическая непараметрическая регрессия для переменной двоичного ответа). Немыслимые всего несколько лет назад методы непараметрического регрессионного анализа стали практичными благодаря достижениям в области статистики и вычислений, и теперь они представляют собой серьезную альтернативу более традиционному параметрическому регрессионному моделированию.

Процедура аппроксимации обычно называется ***сглаживанием***. По существу эта аппроксимация функции отклика Y может быть выполнена двумя способами. Довольно часто используется параметрический подход, заключающийся в предположении, что функция отклика Y имеет некоторую предписанную функциональную форму, например, это прямая линия с неизвестными свободным членом и наклоном. Альтернативой этому может служить попытка оценить Y непараметрическим образом, без указания конкретного ее вида. Первый подход к анализу регрессионной зависимости называется параметрическим, поскольку предполагается, что вид функции полностью описывается конечным набором параметров. Типичный пример параметрической модели представляет собой полиномиальное уравнение регрессии, когда параметрами являются коэффициенты при неизвестных. Однако при параметрическом подходе молчаливо предполагается, что кривая может быть представлена в терминах параметрической модели, или, по крайней мере, имеется уверенность в том, что ошибка аппроксимации для наилучшего параметрического приближения пренебрежимо мала. Наоборот, в непараметрической модели регрессионной зависимости не производится проектирования данных в "прокрустово ложе" фиксированной параметризации. Предварительное задание параметрической модели может оказаться слишком ограничительным или чересчур малой размерности для аппроксимации непредвиденных характеристик, в то время как непараметрическое сглаживание предоставляет гибкие средства анализа неизвестных регрессионных зависимостей. Непараметрический подход приводит, таким образом, к гибкому функциональному виду кривой регрессии.

Определение

В непараметрической регрессии у нас есть случайные величины X и Y и предположим следующие отношения:

$E[Y|\ X = x\ ] = m(x)$,

где $m(x)$ - некоторая детерминированная функция. Линейная регрессия - это ограниченный случай непараметрической регрессии, когда $m(x)$ предполагается аффинным. Некоторые авторы используют более сильное предположение об аддитивном шуме: $Y = m(x) + U$,

где случайная величина U является "шумовым членом" со средним 0. Без предположения, что m принадлежит конкретному параметрическому семейству функций, поэтому получить несмещенную оценку для, однако большинство оценок являются согласованными в подходящих условиях.

Список универсальных алгоритмов непараметрической регрессии

Это неполный список алгоритмов, подходящих для задач непараметрической регрессии.

- ближайшие соседи
- деревья регрессии
- регрессия ядра
- локальная регрессия
- многомерные сплайны адаптивной регрессии
- нейронные сети
- опорная векторная регрессия
- сглаживающие шлицы

Регрессия гауссовского процесса или кригинг

В регрессии гауссовского процесса, также известной как кригинг, для кривой регрессии предполагается гауссовский априор. Предполагается, что ошибки имеют многомерное нормальное распределение, а кривая регрессии оценивается по ее апостериорной моде. Гауссовский априор может зависеть от неизвестных гиперпараметров, которые обычно
оцениваются эмпирическим методом Байеса. Гиперпараме тры обычно определяют предварительное ядро ковариации. В случае, если ядро также должно быть выведено непараметрическим образом из данных, можно использовать критический фильтр

Сглаживающие сплайны интерпретируются как апостериорная мода регрессии гауссовского процесса.

Регрессия ядра

Регрессия ядра оценивает непрерывную зависимую переменную по ограниченному набору точек данных путем свертки местоположений точек данных с помощью функции ядра - грубо говоря, функция ядра указывает, как «размыть» влияние точек данных, чтобы их значения могли быть используется для прогнозирования стоимости для ближайших местоположений.

Деревья регрессии

Алгоритмы обучения дерева решений могут применяться, чтобы научиться предсказывать зависимую переменную на основе данных. Хотя исходная формулировка дерева классификации и регрессии (CART) применялась только для прогнозирования одномерных данных, эту структуру можно использовать для прогнозирования многомерных данных, включая временные ряды.