Статистическая наука
1996, т. 11, № 2, 89–121

Гибкое сглаживание с помощью B-шлицев и штрафы

Пол Х.К. Эйлерс и Брайан Д. Маркс

Абстрактный. B-сплайны привлекательны для непараметрического моделирования, но выбор оптимального количества и положения узлов - сложная задача. Можно использовать равноудаленные узлы, но их небольшое дискретное количество почти всегда. минимумы только ограниченный контроль над гладкостью и посадкой. Предлагаем использовать относительно большое количество узлов и штраф за разность коэффициентов смежных B-шлицев. Мы показываем связи со знакомым сплайном штрафа от интеграла от квадрата второй производной. Краткий обзор B-представлены шлицы, их конструкция и штрафная вероятность. Мы обсудить свойства штрафных B-сплайнов и предложить различные критерии для выбор оптимального параметра штрафа. Непараметрическая логистическая пере-Грессия, оценка плотности и сглаживание диаграммы рассеяния используются как экс-обильные. Приведены некоторые детали расчетов.

Ключевые слова и фразы: обобщенные линейные модели, сглаживание, не-параметрические модели, сплайны, оценка плотности.

1. ВВЕДЕНИЕ

Вряд ли можно сомневаться в том, что сглаживание видное место в статистике сегодня. Многие документы и появился ряд книг (Silverman, 1986; Юбэнк, 1988; Хасти и Тибширани, 1990; Härdle, 1990; Вахба, 1990 г .; Палочка и Джонс, 1993; Зеленый и Сильверман, 1994). Есть несколько причин для эта популярность: многие наборы данных слишком «богаты», чтобы быть полностью смоделированным с помощью параметрических моделей; графический презентация становится все более важной приятный и простой в использовании; и поисковый анализ данные стали более распространенными.

На самом деле, название непараметрическое не всегда хорошо выбран. Это может относиться к сглаживателям ядра и статистику бега, но сглаживающие сплайны де-прописаны по параметрам, хотя их количество может быть большим. Возможно, лучше было бы поговорить о «сверхпара-метрические »методы или« анонимные »модели; Па-параметры не имеют научного толкования.

Пол Х.К. Эйлерс - руководитель отдела в ком-размещение секции DCMR Milieudienst Rijnmond, s-Gravelandseweg 565, 3119XT Schiedam, Нидерланды (электронная почта: paul@dcmr.nl). Брайан Д. Маркс доцент кафедры экспериментальной Статистика, Университет штата Луизиана, Батон-Руж, LA 70803-5606 (электронная почта: brian@stat.lsu.edu).

Существует несколько усовершенствований текущей статистики: тики, как сглаживание ядра (Silverman, 1986; Härdle, 1990) и LOWESS (Кливленд, 1979). Сплайны бывают нескольких разновидностей: сглаживающие шлицы, регрессионные сплайны (Eubank, 1988) и B-сплайны (де Бур, 1978; Диркс, 1993). С таким количеством технологий доступные методы, почему мы должны предлагать новый один? Мы считаем, что сочетание B-шлицев и штрафы за разницу (по расчетному коэффициенту cients), которые мы называем P-сплайнами, имеет очень привлекательные характеристики. P-шлицы не имеют граничных эффектов, они являются прямым расширением (обобщенных) лин-модели регрессии уха, сохранить моменты (означает, дисперсии) данных и имеют полиномиальную кривую подходит как пределы. Расчеты, в том числе для перекрестной проверки, относительно недороги и легко встраивается в стандартное программное обеспечение.

B-сплайны строятся из полиномиальных частей, соединились при определенных значениях x, узлы. Однажды даны узлы, легко вычислить B-сплайны рекурсивно, для любой желаемой степени поли-номинальный; см. de Boor (1977, 1978), Cox (1981) или Диркс (1993). Выбор узлов был предмет большого исследования: слишком много узлов приводят к переоснащение данных, слишком мало узлов приводит к не-дерфиттинг. Некоторые авторы предложили автоматическое схемы оптимизации количества и положения узлы (Friedman, Silverman, 1989; Куперберг и Стоун, 1991, 1992). Это сложно культовая числовая задача и, насколько нам известно, нет привлекательная универсальная схема существует.

Другой трек был выбран О'Салливаном (1986, 1988 г.). Он предложил использовать относительно большое число Бер узлов. Чтобы предотвратить переоснащение, наложен штраф на вторая производная ограничивает гибкость подобранная кривая, аналогичная штрафу, введенному за сглаживающие сплайны Райншем (1967) и стали стандартом в большей части сплайновой литературы. ture; см., например, Eubank (1988), Wahba (1990) и Грин и Сильверман (1994). В этой статье мы упростить и обобщить подход О'Салливана, таким образом, чтобы его можно было применять в любых условиях. текст, в котором полезна регрессия на B-сплайнах. Только небольшие модификации уравнений регрессии нужно.

Основная идея состоит в том, чтобы не использовать интеграл от квадрат высшей производной подобранной кривой в штраф, но вместо этого использовать простую разницу штраф на сами коэффициенты смежных B-шлицы. Мы показываем, что оба подхода очень аналогично для различий второго порядка. В некоторых приложениях катионы, однако может быть полезно использовать различия меньшего или более высокого порядка в размере штрафа. С участием в нашем подходе просто включить штраф в размере любой порядок в (обобщенных) уравнениях регрессии.

Основная проблема любой техники сглаживания: выбор оптимального количества сглаживания, в В нашем случае это оптимальный вес штрафа. Мы используем перекрестная проверка и информационный критерий Акаике - Рион (AIC). В последнем эффективном измерении то есть эффективное количество параметров модель играет решающую роль. Мы следуем за Хасти и Тибширани (1990) в использовании следа более гладкого матрица как эффективное измерение. Потому что мы используем стандартные методы регрессии, это количество может легко вычислить. Мы считаем след очень полезным сравнить эффективное сглаживание для разное количество узлов, разная степень B-сплайны и разные порядки штрафов.

Исследуем сохранение моментов разный порядок, в зависимости от степени B-шлицы и порядок отличий в штраф. Чтобы проиллюстрировать использование P-сплайнов, мы представить в качестве приложений: сглаживание диаграммы рассеяния; моделирование кривых доза – реакция; а также оценка плотности.

2. B-СПЛАЙН В ОДНОМ СЛУЧАЕ

Не все читатели будут знакомы с B-шлицами. Основные ссылки: de Boor (1978) и Dierckx. (1993), но, чтобы проиллюстрировать основную простоту идеи, здесь мы объясняем некоторые важные предпосылки. B-сплайн состоит из полиномиальных частей, соединенных особым образом. Очень простой пример показан на Слева на Рисунке 1 (а): один B-сплайн степени 1. Он состоит из двух линейных частей; одна штука от x_1 до x_2, другой от x_2 до x_3. Узлы - x_1, x_2 и x_3. Слева от x_1 и справа от x_3 этот B-сплайн равно нулю. В правой части рисунка 1 (а) еще три Показаны B-сплайны степени 1: каждый основан на три узла. Конечно, мы можем построить и большие набор B-сплайнов, как нам нравится, введя больше узлы.

В левой части рисунка 1 (б) B-сплайн показана степень 2. Он состоит из трех квадратичных кусочки, соединенные двумя узлами. В точках соединения нет совпадают только ординаты полиномиальных частей, но и их первые производные равны (но не их вторые производные). B-сплайн основан на четыре соседних узла: x_1 x_4. В правой части На рис. 1 (б) изображены еще три B-сплайна степени 2. показано.

Обратите внимание, что B-шлицы перекрывают друг друга. B-сплайны первой степени перекрываются двумя соседями, B-сплайны второй степени с четырьмя соседями и так на. Конечно, крайний левый и крайний правый шлицы иметь меньше перекрытий. При заданном x две первой степени (или три второй степени) B-сплайны отличны от нуля.

Эти примеры иллюстрируют общие свойства B-сплайна степени q:

- состоит из q + 1 полиномиальных частей, каждый из степень q;
- полиномиальные части соединяются в q внутренних узлах;
- в точках стыковки производные до заказа q - 1 непрерывны;
- B-сплайн положителен в области, охватываемой q + 2 узла; везде - ноль;
- кроме границ, перекрывается с 2q полиномиальные части его соседей;
- при заданном x q + 1 B-сплайнов отличны от нуля.

Пусть область от x_{min} до x_{max} разбита на n равных интервалов по n +1 узлов. Каждый интервал будет покрывается q + 1 B-сплайнами степени q. Общая количество узлов для построения B-шлицев будет n + 2q + 1. Количество B-сплайнов в регрессия равна n = n + q. В этом легко убедиться построение графиков, подобных тем, что на рисунке 1.

B-шлицы очень привлекательны в качестве базовых функций для («Непараметрическая») одномерная регрессия. Линейный комбинация (скажем) B-сплайнов третьей степени дает плавная кривая. Как только можно вычислить B-сплайны сами, их применение не сложнее чем полиномиальная регрессия.

Де Бур (1978) дал алгоритм для вычисления B-шлицы любой степени от B-шлицев более низкой степени. Поскольку B-сплайн нулевой степени является константой на один интервал между двумя узлами, легко вычислить

Стр. 3

Рис. 1. Иллюстрации одного изолированного B-сплайна и нескольких перекрывающихся (а) степени 1; (б) степень 2.

Установите B-шлицы любой степени. В этой статье мы используем только эквидистантные узлы, но алгоритм де Бура также работает для любого размещения узлов. Для равноудаленного узлов алгоритм может быть еще более упрощен, так как иллюстрируется небольшой функцией MATLAB в Приложение.

Пусть B_j xq обозначает значение в x j-го B-сплайн степени q для заданной эквидистантной сетки узлы. Подгонка кривой $\hat{y}$ к данным x_i y_i линейный комбинация $\hat{y}x = \sum^{n}_{J=1} {}_J B_J$ x кв. Когда детип B-сплайнов ясен из контекста, или несущественно, мы используем B_j x вместо B_j x q.

Индексирование B-шлицев требует некоторой осторожности, особенно в первую очередь, когда мы собираемся использовать деривативы. B-dexing соединяет B-шлиц с узлом; то есть дает индекс узла, характеризующий положение B-сплайна. Наш выбор - взять крайний левый узел, узел, в котором B-сплайн начинает складываться. ненулевые. На рисунке 1 (а) x_1 - это позиционирование. узел для первого B-шлица. Этот выбор индексации требует, чтобы мы поместили q узлов слева от область x. В следующих формулах для производных tives, точные границы индекса в суммах равны несущественны, поэтому мы их не упомянули.

Де Бур (1978) дает простую формулу для получения Наборы B-шлицев:

$$ч \sum_{дж} a_j B_j xq = \sum_j \quad a_j B_j xq - 1$$

$$(1) \qquad - \sum_j \quad a_{j+1} \; B_{j+1} xq - 1$$

$$= - \sum_j \quad a_{j+1} \; B_j xq - 1$$

где h - расстояние между узлами и a_j = $a_j - a_{j-1}$.

По индукции для второго производная:

$$(2) \qquad ч \; {}_2 \sum_j \; a_j B_j xq = \sum_j \quad {}_{2J} B_J XQ - 2$$

где ${}_2 a_j$ = $a_j = a_j - 2a_{j-1} + a_{j-2}$. Этот факт окажется очень полезным при сравнении непрерывных и дискретные штрафы за шероховатость в следующем разделе.

3. ШТРАФЫ

Рассмотрим регрессию m точек данных x_i y_i на множестве из n B-сплайнов $B_j \cdot$. Наименьшие квадраты об-активная функция минимизации

$$(3) \qquad S = \sum^{m}_{я=1} \{у_я - \sum^{n}_{j=1} a_j B_j x_i\}^2$$

Пусть количество узлов относительно велико, например что подобранная кривая покажет больше вариаций, чем оправдано данными. Чтобы результат был менее гибким - ible, О'Салливан (1986, 1988) ввел штраф на второй производной аппроксимированной кривой, и поэтому сформировали целевую функцию

$$(4) \qquad S = \sum^{m}_{я=1} \{у_я - \sum^{n}_{j=1} a_j B_j x_i\}^2 + \lambda \int^{x_{макс}}_{x_{мин}} \{\sum^{n}_{j=1} a_j B_j x\}^2 dx$$

Интеграл от квадрата второй производной подогнанной функции стало обычным явлением как плавный штрафа, так как основополагающая работа по сглаживанию шлицы Райнша (1967). Нет ничего особенного рассказ о второй производной; на самом деле ниже или также могут использоваться более высокие порядки. В контексте сглаживающих сплайнов первая производная приводит к простые уравнения и кусочно-линейная аппроксимация, в то время как высшие производные приводят к довольно сложным математическим математика, системы уравнений с широким диапазоном частот, и очень гладкая посадка.

Мы предлагаем основывать штраф на (более высокой степени) конечные разности коэффициентов соседних B-шлицы:

$$(5) \qquad S = \sum^{m}_{я=1} \{у_я - \sum^{n}_{j=1} a_j B_j x_i\}^2 + \lambda \sum^{n}_{j=k+1} {}_k a_j{}^2$$

Стр. 4

Такой подход снижает размерность проблема к n, количество B-шлицев, а не м, количество наблюдений, со сглаживанием шлицы. У нас еще есть параметр λ для непрерывных контроль плавности посадки. Различия штраф - хорошее дискретное приближение к внутреннему интегрированный квадрат k-й производной. Что еще важно: с этим штрафом моменты данных являются сохраняющимися и полиномиальными регрессионными моделями. cur как пределы для больших значений λ. См. Раздел 5 для подробности.

Ниже мы покажем, что существует очень сильная связь между штрафом по дифф-коэффициентов B-сплайна и коэффициентов О'Салливана выбор штрафа по второй производной от встроенная функция. Тем не менее, наш штраф может быть применен-обрабатывается механически для любого порядка различий (см. реализацию в Приложении).

Штрафы за разницу имеют долгую историю, вернемся, по крайней мере, к Уиттекеру (1923); недавняя заявка-Все это было описано Грином и Янделлом. (1985) и Эйлерс (1989, 1991a, b, 1995).

lap, когда j равно k - 1, k или k + 1. Таким образом, мы имеем

$$(9) \qquad h^2 P = \lambda \int^{x_{макс}}_{x_{мин}} [\{\sum_j \; {}_2 a_j B_j x \, 1\}^2 + 2 \sum_j \; {}_{2J} \; {}_2 a_{j-1} \cdot B_j x \, 1 B_{j-1} x \, 1] \, dx$$

или же

$$(10) \qquad h^2 P = \lambda \sum_j \; {}_{2J} \; {}_2 \int^{x_{макс}}_{x_{мин}} B^2_j x \, 1 \, dx + 2\lambda \sum_j \; {}_{2J} \; {}_2 a_{j-1} \cdot \int^{x_{макс}}_{x_{мин}} B_j x \, 1 B_{j-1} x \, 1 \, dx$$

что можно записать как

$$(11) \qquad h^2 P = \lambda \{c_1 \sum_j \; {}_{2J} \; {}^2 + c_2 \sum_j \; {}_{2J} \; {}_2 a_{j-1}\}$$

где c_1 и c_2 - константы для заданного (равноудаленного)

Штраф за разницу легко ввести в уравнения регрессий. Это дает возможность поэкспериментируйте с разным порядком различий. В некоторых случаях полезно работать даже с четвертый или более высокий порядок. Это связано с тем, что что для высоких значений λ аппроксимированная кривая приближается к параметрическая (полиномиальная) модель, как будет показано ниже. ниже.

O'Салливан (1986, 1988) использовал третью степень B-шлицы и следующий штраф:

$$h_2 P = \lambda \int_{x_{мин}}^{x_{макс}} \{\sum_j a_j B_j x\ 3\}^2 dx \tag{6}$$

Из производных свойств B-сплайнов следует: минимумы, что

$$h_2 P = \lambda \int_{x_{мин}}^{x_{макс}} \{\sum_j {}_2 a_j B_j x\ 1\}^2 dx \tag{7}$$

Это можно записать как

$$h_2 P = \lambda \int_{x_{мин}}^{x_{макс}} \sum_j \sum_k {}_2 J \ {}_2 a_K \cdot B_j x\ 1 B_k x\ 1\, dx \tag{8}$$

Большинство перекрестных произведений $B_j x\ 1$ и $B_k x\ 1$ исчезают, потому что B-сплайны степени 1 только узлы:

$$c_1 = \int_{x_{мин}}^{x_{макс}} B^2_j x\ 1\, dx$$
$$c_2 = \int_{x_{мин}}^{x_{макс}} B_j x\ 1 B_{j-1} x\ 1\, dx \tag{12}$$

Первый член в (11) эквивалентен нашему второму - Штраф за разницу в порядке, второй член содержит скрещенные произведения соседних вторых разностей. Это приводит к более сложным уравнениям, когда мини-уменьшение штрафной вероятности (уравнения, в которых встречаются семь соседних букв a_j, по сравнению с пятью, если только квадраты секундных разностей возникают в штрафных). Более высокая сложность штрафных уравнений происходит из-за перекрытия B-шлицев. С участием различия более высокого порядка и / или более высокие степени B-шлицы, осложнения быстро растут и затрудняют создание автоматического процедура включения штрафа в вероятность уравнения вытяжки. С применением штрафа за разницу на коэффициенты B-сплайнов эта задача исчезает.

4. НАКАЗАНИЕ ВЕРОЯТНОСТИ

Для сглаживания методом наименьших квадратов мы должны минимизировать S в (5). Система уравнений, вытекающая из минимизацию S можно записать как:

$$B^T y = B^T B + \lambda D^T_k D_k a \tag{13}$$

где D_k - матричное представление различных оператора $_k$, а элементами B являются $b_{ij} = B_j x_i$. При λ = 0 имеем стандартную нормальную

Стр. 5

уравнения линейной регрессии с базисом B-сплайнов. При k = 0 мы имеем частный случай гребневых регрес-сион. При λ> 0 штраф влияет только на главная диагональ и k поддиагоналей (по обе стороны от главная диагональ) системы уравнений. Этот система имеет полосчатую структуру из-за ограниченного Допускается перекрытие B-шлицев. Редко стоит проблема с использованием этой специальной структуры, так как число-число уравнений равно количеству шлицев, которая обычно умеренная (10–20).

В обобщенной линейной модели (GLM) мы во-t создать линейный предиктор $\eta_i = \sum^n_{j=1} b_{ij} a_j$ и a (каноническая) функция зацепления $\eta_i = g\mu_i$, где μ_i - ожидание y_i. Штраф теперь вычитается от логарифма вероятности образования штрафных санкций функция правдоподобия

$$L = ly\ a - \frac{\lambda}{2} \sum^n_{j=k+1} {}_k a_j{}^2 \tag{14}$$

Оптимизация L приводит к следующей системе уравнений:

$$B^T y - \mu = \lambda D^T_k D_k a \tag{15}$$

Они решаются как обычно с итеративно взвешенными линейные регрессии с системой

$$B^T \tilde{W}y - \tilde{\mu} + B^T \tilde{W} B \tilde{a}$$
$$= B^T \tilde{W} B + \lambda D^T_k D_k a \tag{16}$$

где $\tilde{a}$ и $\tilde{\mu}$ - текущие приближения к решение, а $\tilde{W}$ - диагональная матрица весов

$$w_{ii} = \frac{1}{v_i (\partial\mu \partial\eta_i)^2} \tag{17}$$

где v_i - дисперсия y_i при заданном μ_i. Единственный точно соответствуют данным (de Boor, 1977). То же самое справедливо для P-сплайнов, если порядок штрафа k + 1 или выше при любом значении λ. Чтобы увидеть, что это верно, возьмем случай штрафа первого порядка и соответствие данным y, которые являются постоянными (многочлен от степень 0). Потому что $\sum^n_{J=1} B_J x = c$, получаем, что $\sum^n_{J=1} B_J x_я = 0$ для всех x. Тогда из взаимосвязь между различиями и производными финансовыми инструментами в (1) все a_j равны нулю и, следовательно, that $\sum^n_{j=2} a_j = 0$. Следовательно, штраф не действует, и посадка такая же, как и для нештатных B-шлицев. Этот рассуждения легко могут быть расширены индукцией до данные с линейной зависимостью между x и y, и штраф за разницу во втором порядке.

P-шлицы могут сохранять моменты данных. Для линейная модель с P-сплайнами степени k + 1 и а штраф порядка k + 1 или выше, выполняется

$$\sum^n_{я=1} x_к y_я = \sum^n_{я=1} x_k \hat{y}_i \tag{18}$$

для всех значений λ, где $\hat{y}_i = \sum^n_{j=1} b_{ij} \hat{a}_j$ подходят-ted values. Для GLM с каноническими ссылками выполняется что

$$\sum^n_{я=1} x_к y_я = \sum^n_{я=1} x_k \hat{\mu}_i \tag{19}$$

Это свойство особенно полезно в контексте сглаживание плотности: среднее значение и дисперсия расчетная плотность будет равна среднему значению и дисперсии данных для любого сглаживания. Это преимущество по сравнению с сглаживателями ядра: эти увеличивать дисперсию более сильными отличие от стандартной процедуры установки GLM (McCullagh and Nelder, 1989), с B-сплайны как регрессоры, является модификацией $B^T \tilde{W} B$ через $\lambda D_k^T D_k$ (которая сама по себе является постоянной при фиксированном λ) при каждая итерация.

5. СВОЙСТВА P-ШЛИФОВ.

П-шлицы обладают рядом полезных свойств, частично унаследован от B-шлицев. Даем короткое обзор, с несколько неформальными доказательствами.

Во-первых: P-шлицы не показывают граничных эф-влияет, как и многие другие типы сглаживающих веществ. Этим мы имеем в виду распространение подобранной кривой или плотность вне (физического) домена данных, как правило, союзник сопровождается изгибом к нулю. В Сек-В разделе 8 этот аспект рассмотрен более подробно в контекст сглаживания плотности.

P-шлицы могут точно соответствовать полиномиальным данным. Пусть данные $x_i\ y_i$ быть данным. Если y_i - многочлен от x от степени k, то B-сплайны степени k и выше будут сглаживание.

Предел подгонки P-шлицев с сильным сглаживанием является многочленом. Для больших значений λ и штрафа порядка k, подобранный ряд будет приближаться к полиномиальной модели степени k - 1, если степень B-шлицев равно или больше k. Еще раз, родственник зависимости между производными от посадки B-шлица и разности коэффициентов, как в (1) и (2), являются ключ. Возьмем, к примеру, разность второго порядка штраф: когда λ велико, $\sum_{j=3}^{n} (\Delta^2 a_j)^2$ должно быть очень около нуля. Таким образом, каждое из вторых отличий имеет быть близким к нулю, и, следовательно, вторая производная от соответствие должно быть везде близким к нулю. С учетом эти очень полезные результаты, кажется, что B-сплайны а штрафы за разницу - идеальный брак.

Важно сосредоточиться на линеаризованном сглаженном проблема, которая решается на каждой итерации, потому что воспользуемся свойствами сглаживающей ма-трицы. Из (16) следует для шляпной матрицы H:

$$H = B(B^T \tilde{W} B + \lambda D_k^T D_k)^{-1} B^T \tilde{W} \tag{20}$$

След H будет приближаться к k с увеличением λ. A Доказательство состоит в следующем. Позволять

$$Q_B = B^T \tilde{W} B \text{ и } Q_\lambda = \lambda D^T D \tag{21}$$

Запишите trH как

$$\begin{aligned} trH &= tr\{(Q_B + Q_\lambda)^{-1} Q_B\} \\ &= tr\{Q_B^{1/2}(Q_B + Q_\lambda)^{-1} Q_B^{1/2}\} \\ &= tr\{(I + Q_B^{-1/2} Q_\lambda Q_B^{-1/2})^{-1}\} \end{aligned} \tag{22}$$

Это можно записать как

$$trH = tr(I + \lambda L)^{-1} = \sum_{j=1}^{n} \frac{1}{1 + \lambda\gamma_j} \tag{23}$$

где

$$L = Q_B^{-1/2} Q_\lambda Q_B^{-1/2} \tag{24}$$

а γ_j при j = 1 n - собственные значения оператора L. Поскольку k собственных значений Q_λ равны нулю, L имеет k нулевых собственные значения. Когда λ велико, только k членов с $\gamma_j = 0$ вносят вклад в крайний левый член, а таким образом, к следу H. Следовательно, trH приближается к k для больших λ.

6. ОПТИМАЛЬНОЕ Сглаживание, AIC И ПЕРЕКРЕСТНАЯ ПРОВЕРКА

Теперь, когда мы можем легко влиять на плавность аппроксимированной кривой с λ, нам нужно каким-то образом выбрать «оптимальное» значение для него. Предлагаем использовать Информационный критерий Акаике (AIC).

Основная идея AIC - исправить журнал регистрации. вероятность подобранной модели для эффективного числа параметров. Обширное обсуждение и применение катионы можно найти в Сакамото, Исигуро и Китагава (1986). Вместо логарифмической вероятности отклонение легче использовать. Определение AIC: эквивалентно

$$AIC\lambda = devy\ a\ \lambda + 2 * dima\ \lambda \tag{25}$$

где dima λ - (эффективный) размер вектор параметров, a, и devy a λ - это отклонение.

Поскольку trAB = trBA (для согласованных матриц), с вычислительной точки зрения выгодно использовать

$$\begin{aligned} trH &= tr\{B(B^T WB + \lambda D_k^T D_k)^{-1} B^T W\} \\ &= tr\{B^T WB(B^T WB + \lambda D_k^T D_k)^{-1}\} \end{aligned} \tag{26}$$

Последнее выражение включает только матрицы размером n на n, тогда как H - это матрица размером m на m.

В некоторых GLM масштаб данных известен, как для отсчетов с распределением Пуассона, так и для биномиальные данные; тогда отклонение может быть вычислено напрямую. Для линейных данных оценка дисперсии необходим. Один из подходов - взять дисперсию остатки от $\hat{y}_i$, которые вычисляются, когда $\lambda = 0$, скажем, $\hat{\sigma}_0^2$:

$$\begin{aligned} AIC &= \sum_{i=1}^{m} \frac{(y_i - \hat{\mu}_i)^2}{\hat{\sigma}_0^2} + 2\, trH \\ &\quad - 2m \ln \hat{\sigma}_0 - m \ln 2\pi \end{aligned} \tag{27}$$

Этот выбор дисперсии довольно произвольный, так как это зависит от количества узлов. Альтернативы могут основываться на (обобщенной) перекрестной проверке. Для обычных ни одной перекрестной проверки мы вычисляем

$$CV\lambda = \sum_{i=1}^{m} \left(\frac{y_i - \hat{y}_i}{1 - h_{ii}}\right)^2 \tag{28}$$

где h_{ii} - диагональные элементы шляпы матрица H. Для обобщенной перекрестной проверки (Wahba, 1990), мы вычисляем

$$GCV\lambda = \sum_{i=1}^{m} \frac{(y_i - \hat{y}_i)^2}{(m - \sum_{i=1}^{m} h_{ii})^2} \tag{29}$$

Разница между обеими величинами обычно составляет небольшой. Лучшее λ - это значение, которое минимизирует $CV\lambda$ или $GCV\lambda$. Дисперсия остатков на оп-Timal λ - естественный выбор для использования в качестве оценки σ_0^2 для вычисления $AIC\lambda$. Практично работать с модифицированными версиями $CV\lambda$ и $GCV\lambda$, со значениями, которые можно интерпретировать как оценки стандартное отклонение перекрестной проверки:

$$\begin{aligned} \bar{CV}\lambda &= \sqrt{CV\lambda / m} \\ \bar{GCV}\lambda &= \sqrt{m\, GCV\lambda} \end{aligned} \tag{30}$$

Вычислить отклонение несложно, но как определить эффективный размер подходят ли наши P-шлицы? Мы находим решение в Hastie и Тибширани (1990). Они обсуждают эффективный размер- линейных сглаживателей и предлагаем использовать след более гладкой матрицы в качестве приближения. В нашем случае это означает dima = trH. Обратите внимание, что trH = n при λ = 0, как в (неособом) стандарте линейная регрессия.

Два члена в AICλ представляют отклонение и след более гладкой матрицы. Последний член, скажем, Tλ = trHλ, представляет интерес сам по себе, потому что его можно интерпретировать как эффективный размер вид подобранной кривой.

Tλ полезно для сравнения подгонок для разных чисел. узлов и порядков штрафов, а λ может варьируются в широком диапазоне значений и не имеют четких интуитивное обращение. Мы покажем на примере ниже

Стр.7

Таблица 1
Значения нескольких диагностик для данных удара мотоцикла, для нескольких значений λ

λ	0,001	0,01	0,1	0,2	0,5	1	2	5	10
резюме	24,77	24.02	23,52	23,37	23,26	23,38	23,90	25,50	27,49
GCV	25,32	24,93	24,17	23,94	23,74	23,81	24,28	25,87	27,85
AIC	159,6	156,2	149,0	146,7	144,7	145,4	150,6	169,1	194,3
trH	21,2	19,4	15,13	13,6	11,7	10,4	9.2	7,7	6,8

что график зависимости AIC от T является полезным диагностическим инструмент.

В случае P-шлицев максимальное значение, которое Tλ может достигать, равно количеству B-сплайнов (при λ = 0). Фактический максимум зависит от количество и распределение точек данных. В минимальное значение Tλ происходит, когда λ стремится к бесконечности. ity; он равен разнице штрафа. Это согласуется с тем, что при больших значениях λ подгонка P-сплайнов приближается к полиному де- греческий к - 1.

7. ПРИЛОЖЕНИЯ К ОБЩИМ ЛИНЕЙНОЕ МОДЕЛИРОВАНИЕ

В этом разделе мы применяем P-сплайны к ряду ситуации непараметрического моделирования с нормальным а также ненормальные данные.

Сначала рассмотрим проблему с аддитивными ошибками. Сильверман (1985) использовал мотоциклетный шлем im- данные пакта для иллюстрации сглаживания диаграммы рассеяния со шлицами; данные можно найти в Härdle (1990) и (также на дискете) в Hand et al. (1994). В данные дают ускорение головы в единицах g, при разных время после удара в моделируемых авариях. Мы гладкие с B-шлицами степени 3 и второ- штраф за заказ. Выбранные узлы делят область x (0–60) на 20 интервалов одинаковой ширины. Когда варьируем λ на приближенно геометрической сетке, получить результаты в таблице 1, где $\hat\sigma_0$ вычисляется от GCVλ при оптимальном значении λ. На оп- временное значение λ, определяемое GCV, мы получаем результаты, как показано на рисунке 2.

Интересно отметить, что объем работы исследовать несколько значений λ во многом не зависит. Влияние количества точек данных при использовании GCV. Решаемая система:

$$(31) \qquad B^T B + \lambda D_k^T D_k a = B^T y$$

Сумма квадратов равна

$$(32) \qquad S = |y - Ba|^2 = y^T y - 2a^T B^T y + a^T B^T Ba$$

Таким образом, $B^T B$ и $B^T y$ должны вычисляться только один раз. Матрица шляпы H имеет размер m на m, но для ее следа мы нашел выражение в (26), которое включает только $B^T B$ и $D_k^T D_k$. Так что нам не нужны исходные данные для перекрестная проверка при любом значении λ.

Наш второй пример касается логистической регрессии. Модель

$$(33) \qquad \ln\left(\frac{p_i}{1 - p_i}\right) = \eta_i = \sum_{j=1}^{n} a_j B_j(x_i)$$

Наблюдения представляют собой троек $x_i\ t_i\ y_i$, где t_i - количество исследуемых лиц в дозе x_i, y_i - количество «успехов». Мы предполагаем, что y_i имеет биномиальное распределение с вероятностью p_i и t_i испытаний. Ожидаемое значение y_i равно $t_i p_i$ и дисперсия равна $t_i p_i (1 - p_i)$.

На рисунке 3 показаны данные Эшфорда и Уокера. (1972) о количестве трипаносомных организмов убивали при разных дозах определенного яда. Данные показаны точки и две подогнанные кривые. Для толстых линия кривой λ = 1 и AIC = 134; это значение λ равно оптимально для выбранных B-сплайнов степени 3 и а штраф порядка 2. Тонкая кривая показывает подходит для $λ = 10^8$ (AIC = 278). Со вторым порядком штраф, это, по сути, логистическая подгонка.

На рис.4 показаны зависимости AICλ от Tλ при различные значения k, порядок штрафа. Мы найти, что k = 3 может дать более низкое значение AIC (для λ = 5, AIC = 118). При k = 4 получаем, что очень допускается высокое значение λ; то AIC = 114, вряд ли отличается от минимально возможного значения (11.1). А большое значение λ со штрафом четвертого порядка означает что фактически аппроксимированная кривая для η является кривой третьего порядка полином. Предел посадки с P-шлицами таким образом указывает на кубическую логистическую подгонку как на хороший параметрический модель. Здесь мы видели приложение, в котором штраф четвертого порядка полезен.

Наш третий пример - это временной ряд значений y_i, который мы будем моделировать с помощью распределения Пуассона с плавно меняющимся ожиданием:

$$(34) \qquad \ln \mu_i = \eta_i = \sum_{j=1}^{n} a_j B_j(x_i)$$

В этом частном случае x_i равноудалены, но это несущественно. На рис. 5 показаны номера дис- астры на британских угольных шахтах в 1850–1850 гг. 1962 г., как представлено в (Diggle and Marron, 1988). Счетчики отображаются в виде узких вертикальных полос, линия - это подогнанный тренд. Количество интервалов 20 B-шлицы имеют степень 3 и порядок Штраф равен 2. Оптимальное значение λ было найдено. на приблизительно геометрической сетке 1, 2, 5, 10 и

Стр. 8

Рис. 2. Данные о ударе мотоциклетного шлема: оптимальная посадка с B-шлицами третьей степени, штраф второго порядка и λ = 05.

Рис. 3. Непараметрическая логистическая регрессия данных трипаносом: P-сплайны 3-го порядка с 13 узлами, штраф за разность 2-го порядка λ = 1 и AIC = 134 (жирная линия); тонкая линия фактически соответствует логистической подгонке λ = 10 8 и AIC = 278.

Рис. 4. Зависимость AICλ от эффективной размерности Tλ для нескольких порядков штрафа k.

скоро. Минимум AIC (126,0) был найден для λ = 1000.

Исходные данные об авариях на угледобывающих предприятиях приблизительно были даты, когда они произошли. Таким образом, данные, которые мы здесь используем, на самом деле представляют собой гистограмму с корзинами шириной один год. С событиями в масштабе времени кажется естественным сглаживать подсчеты по интервалам, но та же идея применима к любой форме гистограммы (подсчет бункеров) или сглаживание плотности. Это уже было отмечены Дигглом и Марроном (1988). В следующую секунду- Мы подробно рассмотрим сглаживание плотности с P-шлицами.

8. РАЗГЛАЖИВАНИЕ ПЛОТНОСТИ

В предыдущем разделе мы отметили, что время Число отсчетов - это просто гистограмма на оси времени. Любая другая гистограмма может быть сглажена тем же

Стр.9

Рис. 5. Количество тяжелых аварий на угольных шахтах Великобритании: количество за год показано вертикальными линиями; подобранный тренд ожидания Распределение Пуассона; B-сплайны степени 3 штрафа порядка 3 20 интервалов между 1850 и 1970 годами λ = 1000 и AIC = 1260.

способ. Однако, по нашему опыту, эта идея многим коллегам трудно проглотить. Они видят построение частотной гистограммы как не- допустимая дискретизация данных и как предварительная Люд к катастрофе. Возможно, это чувство проистекает из общеизвестный факт, что максимальная вероятность хронирование гистограмм приводит к патологическим результатам, а именно, дельта-функции при наблюдениях (Скотт, 1992). Однако, если мы оптимизируем штрафную вероятность капота, мы приходим к стабильным и очень полезным результатам, так как мы покажем ниже.

Пусть y_i, i = 1 m, - гистограмма. Пусть ори- джин x выбирается таким образом, чтобы средние точки бинов $x_i = ih$; таким образом y_i - это количество необработанных наблюдения с $x_i - h/2 \leq x < x_i + h/2$. Если p_i равно вероятность найти необработанное наблюдение в ячейке i, тогда вероятность данной гистограммы пропорциональна к полиномиальному правдоподобию $\prod_{я=1}^{m} p_я^{y_я}$. Эквив- (см. Епископ, Финберг и Голландия, 1975 г., Глава 13), можно работать с вероятностью m Распределения Пуассона с ожиданиями $\mu_i = p_i y_+$, где $y_+ = \sum_{я=1}^{м} y_я$.

Чтобы сгладить гистограмму, мы снова используем общий линейная модель с канонической лог-связью (которая гарантирует положительный μ):

(35) $$\ln \mu_i = \eta_i = \sum_{j=1}^{n} a_{кj} B_j x_i$$

и построить вероятность оштрафованного журнала

(36) $$L = \sum_{я=1}^{m} y_i \ln \mu_i - \sum_{я=1}^{m} \mu_i - \lambda \sum_{j=k+1}^{n} \frac{\Delta^k a_j^2}{2}$$

с па подходящим (т. е. относительно большим) количеством узлы для B-шлицев. Штрафная вероятность уравнения следуют из минимизации L:

(37) $$\sum_{я=1}^{m} y_i - \mu_i B_j x_i = \lambda \sum_{л=к+1}^{n} d_{jl} a_l$$

Эти уравнения решаются с итеративным повторением. взвешенная регрессия, как описано в разделе 4.

Пусть теперь h - ширина ячеек гистограммы. togram уменьшите до очень маленького значения. Если сырой данные даны с бесконечной точностью, мы даже- в итоге получается ситуация, в которой каждая ячейка гистограмма содержит не более одного наблюдения. В другом словами, у нас очень большое количество (m) ячеек, из который y_+ равны 1, а все остальные 0. Пусть I - множество индексы ячеек, для которых $y_i = 1$. Тогда

(38) $$\sum_{я=1}^{m} y_i B_j x_i = \sum_{i \in I} B_j x_i$$

Если исходные наблюдения u_t для t = 1 r, с $r = y_+$, то мы можем написать

(39) $$\sum_{i \in I} B_j x_i = \sum_{t=1}^{n} B_j u_t = B_{+j}$$

и уравнения правдоподобия со штрафом в (37) изменить на

(40) $$B_{+j} - \sum_{я=1}^{m} \mu_i B_j x_i = \lambda \sum_{л=к+1}^{n} d_{jl} a_l$$

Для любого j первый член в левой части (40) можно интерпретировать как «эмпирическую сумму» B-сплайн j, а второй член слева может быть интерпретируется как «ожидаемая сумма» этого B-сплайна для подобранной плотности. При λ = 0 эти члены имеют быть равными друг другу для каждого j.

Обратите внимание, что второй член в левой части уравнения (40) фактически является численным приближением интеграл:

(41) $$\sum_{я=1}^{m} \mu_i B_j x_i / y_+ \approx \int_{x_{мин}}^{x_{макс}} B_j x \exp\{\sum_{l=1}^{n} a_l B_l x\}\, dx$$

Стр.10

Таблица 2
Значение AIC при нескольких значениях лямбда для оценки плотности Old Faithful

λ	0,001	0,01	0,02	0,05	0,1	0,2	0,5	1	10
AIC	50,79	48,21	47,67	47,37	47,70	48,61	50,59	52,81	65,66

Чем меньше h (больше m), тем лучше приложение- приближение. Другими словами: дискретизация требуется только для численного решения интеграла для что, насколько нам известно, не является решением в замкнутой форме существуют. Для практических целей достаточно простой суммы. эффективная, но более сложная схема интеграции возможно. Обратите внимание, что суммы для вычисления B_{+j} in- включать все необработанные наблюдения, но на самом деле в каждом из только q + 1 слагаемых $B_j u_t$ добавить к своим кор- спонсируя B_{+j}.

Необходимые вычисления могут быть выполнены в терминах

домен от 0 до 6 был разбит на 20 интервалов для определения узлов. На рисунке две подходят показаны для λ = 0001 и λ = 005. Последний значение дает минимум AIC, как показано в таблице 2. Мы видим, что из двух четко разделенных горбов правый кажется смесью двух пиков.

Второй пример также взят из (Silverman, 1986). Данные - это длительность периодов психиатрических лечения в исследовании самоубийств. На рисунке 7 показан необработанные данные и оценочная плотность, когда main выбирается от 0 до 1000. Третья степень B-

достаточной статистики B_{+j}: мы видели их
роль в приведенных выше уравнениях штрафного правдоподобия. Но
также отклонение и, следовательно, AIC могут быть вычислены
напрямую:

$$
\begin{aligned}
\mathrm{devy}\ a &= 2 \sum_{я=1}^{m} y_i \ln y_i / \mu_i \\
(42) \qquad &= 2 \sum_{я=1}^{m} y_я \,\text{лн}\, y_я - 2 \sum_{я=1}^{m} y_я \sum_{j=1}^{n} a_к{}_j B_j x_i \\
&= 2 \sum_{я=1}^{m} y_я \,\text{лн}\, y_я - 2 \sum_{j=1}^{n} a_j B_{+j}
\end{aligned}
$$

В крайнем случае, когда y_i равны 0 или
1 член $\sum y_i \ln y_i$ обращается в нуль. В любом случае это
независимо от установленной плотности.

Плотность сглаживания с Р-образными шлицами очень
привлекательно: оценочная плотность положительная и
непрерывный, его можно описать относительно экономно-
в частности, в терминах коэффициентов В-сплайнов,
и это правильная плотность. Мгновения сохранены,
как следует из (19). Это означает, что с третьим -
В-сплайны степени и штраф третьего порядка, среднее
и дисперсия оценочного распределения равны
к исходным данным, независимо от количества
сглаживание; предел для высоких λ является нормальным распределением
бюция.

Устройство сглаживания плотности Р-сплайна не беспокоит
граничные эффекты, например, сглаживание ядра
находятся. Маррон и Рупперт (1994) приводят примеры и
довольно сложное лекарство, основанное на преобразовании-
ции. При использовании Р-образных шлицев нет необходимости в каких-либо сп
essary, но важно указать домен
данные правильно. Мы представим пример:
низкий.

Теперь возьмем в качестве первого примера набор данных из
(Сильверман, 1986). Данные длительностью 107
извержения гейзера Старый Верный. Третья степень
Использовались В-шлицы со штрафом третьего порядка. В
шлицы со штрафом второго порядка. А
довольно большая величина сглаживания (λ = 100)
выделено AIC; подобранная плотность почти экспоненциально
нулевой. Фактически, если рассматривать только домен
от 0 до 500, то λ может стать сколь угодно большим
и получается чистая экспоненциальная плотность. Однако если
выбираем область от −200 до 800, получаем
совершенно другая посадка, как показано на рисунке 8. Расширяя
области мы заставляем оценочную плотность также
покрывают отрицательные значения x, где нет данных
(что означает ноль). Следовательно, он должен
падают к нулю, пропуская пик для малых положений
Активные ценности. Оптимальное значение λ теперь составляет 0,01 и
гораздо более волнообразные результаты, с заметным
более высокое значение AIC. Это прекрасно иллюстрирует, как с
При правильном выборе области плотность Р-сплайна
более гладкая может быть свободна от граничных эффектов, которые
доставляет столько хлопот с сглаживателями ядра.

9. ОБСУЖДЕНИЕ

Мы считаем, что Р-шлицы близки к тому, чтобы быть
идеально ровнее. С их основанием в классической ре-
методы агрессии и обобщенные линейные модели,
их свойства легко проверить и понять.
Моменты данных сохраняются, а ограничивающий
поведение с строгим наказанием четко определено и
дает связь с полиномиальными моделями. Граница-
иные эффекты не возникают, если область данных
правильно указано.

Необходимых вычислений, включая перекрестные
валидации, сопоставимы по размеру с таковыми для
задача регрессии среднего размера. Регресс
контекст делает естественным расширение Р-сплайнов на
полупараметрические модели, в которых дополнительные экс-
происходят плановые переменные. Вычисленное соответствие:
компактно описывается коэффициентами В-
шлицы.

Стр. 11

Рис. 6. Сглаживание плотности длительностей извержений гейзера Old Faithful: гистограмма плотности и аппроксимированные значения плотности; тонкая линия, третий порядок
штраф с λ = 0001AIC = 8405; толстая линия, оптимальная λ = 005, при AIC = 8017; В-шлицы степени 3 с 20 интервалами на
домен с 1 по 6.

Рис. 7. Сглаживание плотности данных о самоубийствах: положительный домен (0–1000); B-сплайны степени 3 штрафа порядка 2 20 интервалов, λ = 100 AIC = 699.

Рис. 8. Сглаживание плотности данных о самоубийствах: область включает отрицательные значения –200–800; B-шлицы 3-й степени штрафа 2-го порядка, 20 интервалов, λ = 001, AIC = 836.

P-шлицы могут быть очень полезны в (обобщенных) ад-
дитивные модели. Для каждого размера B-образный шлиц.
sis и штраф. С n узлов в
каждое основание и размер d, система nd-by-nd
(взвешенные) результаты уравнений регрессии. Подгонка,

итеративное сглаживание для каждого отдельного измерения-
sion, устраняется. Мы сообщили об этом приложении.
в другом месте (Marx and Eilers, 1994, 1996).

Вероятность наказания - предмет с растущим
популярность. Мы уже упоминали о работе

Стр. 12

О'Салливан. В книге Грина и Сильвермана
(1994), многие приложения и ссылки могут быть
нашел. Почти исключительно штрафы определены в
члены квадрата второй производной
подобранная кривая. Обобщения на штрафы по высшему
производные упоминались в литературе,
но, насколько нам известно, практическое применение
очень редкий. Переход от непрерывного штрафа к
дискретный штраф в терминах коэффициентов
B-шлицы сами по себе не впечатляют. Но мы
видели, что это приводит к очень полезным результатам, в то время как
давая механический способ работы с более высоким порядком
штрафы. Моделирование биномиальной доза-реакция
в разделе 7 показали полезность более высокого порядка
штрафы.

Замечательным свойством AIC является то, что его легче
вычислить его для некоторых ненормальных распределений, таких как
пуассоновского и биномиального, чем для нормального распределения
ции. Это так, потому что для этих распределений
известна взаимосвязь между средним и дисперсией.
Мы должны предупредить читателя, что AIC может привести к
сглаживание при чрезмерном рассредоточении данных,
поскольку предполагаемая дисперсия данных может быть
слишком низко. В настоящее время мы исследуем сглаживание
с P-сплайнами и сверхдисперсными распределениями, такими как
отрицательный бином и бета-бином. Также
идеи квазивероятности будут включены.

Особое внимание мы уделили плотности гладкой поверхности.
потому что мы считаем, что в этой области
звенья P-образных шлицев действительно сияют. Традиционно ядро
гладильные машины были популярны в этой области, но они
раздуть дисперсию и иметь проблемы с границами
ряд областей данных; их вычисление обходится дорого
sive, перекрестная проверка тем более, и никто не может
сообщать предполагаемую плотность в компактном виде.

Возможно, сглаживание ядра все еще имеет преимущества
в двух или более измерениях, но кажется, что
P-шлицы также можно использовать для двумерных
сглаживание произведениями Кронекера B-шлицев.
С сеткой, скажем, 10 на 10 узлов и третьего порядка
штраф, система из 130 уравнений приводит к

различной жесткости, но совсем другое дело
оцените веса по данным.

Наконец, мы хотели бы отметить, что P-сплайны образуют
мост между чисто дискретным сглаживающим
lem, как первоначально сформулировано Уиттакером (1923) и
непрерывное сглаживание. B-сплайны нулевой степени - это
постоянная на интервале между двумя узлами, и ноль
в другом месте; у них нет перекрытия. Таким образом, подогнанный
функция дает для каждого интервала значение ко-
эффективный соответствующего B-сплайна.

ПРИЛОЖЕНИЕ: ДЕТАЛИ РАСЧЕТОВ

Здесь мы посмотрим на вычисление B-сплайнов
и производные от штрафа. Мы используем S-PLUS и
MATLAB в качестве примеров языков из-за их
Широкое применение. Также приводим некоторые впечатления от
скорость вычислений.

В линейном случае необходимо решить систему
уравнения

$$(B^T B + \lambda D_k^T D_k)\hat{a} = B^T y \tag{43}$$

и для вычисления $|y - B\hat{a}|^2$ и $\mathrm{tr}\,(B^T B + \lambda D^T D)^{-1} \cdot B^T B$. Нам нужна функция для вычисления B, B-
базовая матрица сплайна. В S-PLUS это простой мат-
ter, так как есть встроенная функция spline.des (), которая
вычисляет (производные) B-сплайнов. Нам только нужно
построить последовательность узлов. Предположим, что
xl - это левая часть x-области, xr - правая, и что
в этом домене есть интервалы ndx. Вычислить
B для данного вектора x на основе B-сплайнов степени
bdeg, мы можем использовать следующую функцию:

```
bspline <- function (x, xl, xr, ndx, bdeg) {
  dx <- (xr - xl) / ndx
  узлы <- seq (xl - bdeg * dx, xr + bdeg * dx, by = dx)
  B <- spline.des (knots, x, bdeg + 1, 0 * x) $ design
  B
}
```

Обратите внимание, что S-PLUS работает с порядком Bполовина полосы пропускания примерно 30. Это легко может обрабатываться на персональном компьютере. Автоматический построение уравнений будет сложнее чем в одном измерении. Первые эксперименты с этим подход выглядит многообещающе; мы сообщим о них в в свое время.

Мы не коснулись многих очевидных и информативных Интересные удлинители к П-образным шлицам. Надежность может можно получить с помощью любой нелинейной схемы переназначения которые можно использовать с регрессионными моделями. Круговой домены можно обрабатывать, оборачивая B-сплайны и штраф за происхождение. Штраф может быть расширенным за счет веса, чтобы обеспечить посадку с неконтролируемыми постоянная скованность. Таким образом будет легко указать сплайны, следуя первоначальному определению де Бура (1977): порядок равен степени плюс 1.

Матрица D_k также легко вычисляется. В единичная матрица размера n на n строится diag (n) и есть встроенная функция diff () для разница это. Короткой петлей приходим к D_k. Вычисления, таким образом, представлены как (при этом порядок наложения штрафа) следует:

```
B <- bspline (x, xl, xr, ndx, bdeg)
D <- diag (ncol (B))
для (k в 1: pord) D <- diff (D)
a <- решить (t (B)% *% B + lambda * t (D)% *% D,
        т (B)% *% y)
```

Стр. 13


```
yhat <- B% *% a
s <- сумма ((y - yhat) ^ 2)
Q <- решить (t (B)% *% B + лямбда * t (D)% *% D)
           # инверсия матрицы
t <- сумма (diag (Q% *% (t (B)% *% B)))
gcv <- s / (строка (B) - t) ^ 2
```

Есть место для оптимизации вычислений выше путем сохранения и повторного использования промежуточных результатов.

MATLAB не имеет встроенной функции для вычисления B-сплайны, поэтому мы должны запрограммировать рекурсии на наши-себя. Начнем с рекуррентного соотношения: дано у де Бура (1978, глава 10):

$$\frac{B_{jk}x}{\tau_{j+k}-t_j} = \frac{x-t_j}{\tau_{j+k-1}-t_j}\,\frac{B_{jk-1}x}{\tau_{j+k-1}-t_j} + \frac{t_{jk}-x}{\tau_{j+k}-t_j}\,\frac{B_{j+1\,k-1}x}{\tau_{j+k}-t_{j+1}} \tag{44}$$

где $B_{jk}x$ в обозначениях де Бура - это наша $B_j\,xk-1$ (де Бур использует порядок 1 для постоянных B-сплайнов, тогда как мы используем степень 0). Использование униформы сетка узлов на расстояниях $dx = x_{max} - x_{min} / n$ значительно упрощает формулы. Если мы определим p = $x - x_{min} / dx$, приходим к следующему повторению формула:

$$B_j\,xk = \frac{\kappa + p - j + 1}{k} B_{j-1}\,xk-1 + \frac{j-p}{k} B_j\,xk-1 \tag{45}$$

Рекурсию можно начать с k = 0, потому что $B_j\,x0 = 1$, когда $j - 1dx < x - x_{min} \leq jd$, и ноль для всех остальных j. Кроме того, $B_j\,xk = 0$ для $j < 0$ и $j > п$. Это приводит к следующей функции:

```
функция B = bspline (x, xl, xr, ndx, bdeg)
  dx = (xr - xl) / ndx;
  т = xl + dx * [-bdeg: ndx-1];
  T = (0 * x + 1) * т;
  X = x * (0 * t + 1);
  P = (X - T) / dx;
  B = (T <= X) & (X <(T + dx));
  r = [2: длина (t) 1];
  для k = 1: bdeg
    B = (P. * B + (k + 1 - P). * B (:, r)) / k;
  конец;
конец;
```

Вычисление D_k немного проще, потому что есть встроенная функция diff (), которая принимает следующий:

```
B = bspline (x, xl, xr, ndx, bdeg);
[mn] = размер (B);
D = diff (глаз (п), порд);
a = (B '* B + лямбда * D' * D) \ (B '* y);
yhat = B * a;
Q = inv (B '* B + лямбда * D' * D);
s = сумма ((y - yhat). ^ 2)
t = сумма (diag (Q * (B '* B)));
gcv = s / (m - t) ^ 2;
```

Формулы для уравнения штрафного правдоподобия В предложениях описывается, как включить штраф, когда у каждого есть доступ ко всем отдельным этапам повторного Грессивные вычисления. Если это не так, данные может помочь увеличение. Вместо того, чтобы работать с матрицы B B-сплайнах регрессор и $D_к$ из наказание по отдельности и объединение их внутреннего продукта ucts, расширенные данные могут быть построены следующим образом:

$$[y0] \approx \begin{bmatrix} B \\ \sqrt{\lambda}D_k \end{bmatrix} \tag{46}$$

где ≈ означает регресс левого вектора в правой матрице. Для линейных задач это достаточно сделать это только один раз. В общем линейные модели, необходимо увеличение данных заново в каждой из итераций с взвешенными линейными регрессии.

Мы протестировали приведенные выше фрагменты программы на ПК. с процессором Pentium 75 МГц, с S-PLUS 3.3 и MATLAB 4.2, оба работают под Windows для рабочих групп. Данные были взяты из эксперимент с мотоциклетным шлемом, как показано на рис. ure 2. Имеется 133 точки данных, и мы использовали 20 интервалы на x-области. S-PLUS ушло около 0,9 во-вторых, Matlab около 0,2 секунды (для одного значения λ). Это время может быть уменьшено до 0,6 секунды и 0,1 секунды. во-вторых, соответственно, сохраняя и повторно используя некоторые входящие промежуточные результаты ($B^т B$ и обратное $B^т B + \lambda D^т_k D_k$).

Функции для обобщенного линейного оценивания могут быть полученным от первого автора. Мы готовимся представление в Statlib.

БЛАГОДАРНОСТИ

Наше первоначальное исследование, представленное в Eilers и Маркс (1992) не указывал на то, что работа (O'Sullivan, 1986, 1988) неявно использовала модифицированный штраф за разность второго порядка. Мы благодарен профессору Вахбе за привлечение внимания к этой связи и нашему надзору. Мы также благодарим

параметр для порядка разницы. Conse-
соответственно, в MATLAB вычисления выглядят как

анонимному рефери за множество предложений улучшить нашу презентацию.

Стр. 14

РЕКОМЕНДАЦИИ

Эшфорд, Р. и Уокер, П. Дж. (1972). Количественный анализ отклика sis для смеси популяций. Биометрия 28 981–988.

Бишоп, YMM, Fienberg, SE и Холланд, PW (1975). Дискретный многомерный анализ: теория и практика. Массачусетский технологический институт Нажмите.

Кливленд, WS (1979). Надежная локально взвешенная регрессия и сглаживающие графики рассеяния. J. Amer. Статист. Доц. 74 829–836.

Кокс, MG (1981). Практическая аппроксимация сплайнами. В темах в Численный анализ (П. Р. Тернер, ред.). Спрингер, Берлин.

де Бур, К. (1977). Пакет для расчета с B-шлицами. SIAM J. Numer. Анальный. 14 441–472.

де Бур, К. (1978). Практическое руководство по сплайнам. Спрингер, Берлин.

Диркс, П. (1993). Подгонка кривых и поверхностей с помощью шлицев. Кларендон, Оксфорд.

Диггл П. и Маррон Дж. С. (1988). Эквивалентность гладкой Селекторы параметров при оценке плотности и интенсивности. J. Amer. Статист. Доц. 83 793–800.

Эйлерс, PHC (1990). Сглаживание и интерполяция с помощью gen- обобщенные линейные модели. Quaderni di Statistica e Matematica Applications Alle Scienze Economico-Sociali 12 21–32.

Эйлерс, PHC (1991a). Штрафной регресс в действии: оценка вязка загрязняющих роз из среднесуточных значений. Окружающая среда 2 25–48.

Эйлерс, PHC (1991b). Непараметрическая оценка плотности с сгруппированные наблюдения. Статист. Neerlandica 45 255–270.

Эйлерс, PHC (1995). Косвенные наблюдения, составная ссылка модели и штрафованная вероятность. В статистическом моделировании (GUH Seeber et al., Ред.). Спрингер, Нью-Йорк.

Эйлерс, PHC и Маркс, Б.Д. (1992). Обобщенный линейный модели с П-образными шлицами. Прогресс в области GLIM и статистики Моделирование (Л. Фармейр и др., Ред.). Спрингер, Нью-Йорк.

Юбэнк, Р.Л. (1988). Сглаживание сплайнов и непараметрическая перестановка агрессия. Деккер, Нью-Йорк.

Фридман, Дж. И Сильверман, Б.В. (1989). Гибкий парсимо- интеллектуальное сглаживание и аддитивное моделирование (с обсуждением). Технометрика 31 3–39.

Грин, П.Дж. и Сильверман, Б.В. (1994). Непараметрический Re- Грессия и обобщенные линейные модели. Чепмен и Холл, Лондон.

Грин, П.Дж. и Янделл, Б.С. (1985). Полупараметрический обобщенные линейные модели. В обобщенных линейных моделях (Б. Гилкрист и др., Ред.). Спрингер, Нью-Йорк.

Хэнд, ди-джей, Дейли, Ф., Ланн, А.Д., МакКонвей, К.Дж. и Ос- trowski, E. (1994). Справочник малых наборов данных. Глава- Человек и Холл, Лондон.

Хердл, В. (1990). Прикладная непараметрическая регрессия. Cam- мост Univ. Нажмите.

Хасти, Т. и Тибширани, Р. (1990). Обобщенный аддитивный мод- els. Чепмен и Холл, Лондон.

Куперберг, К. и Стоун, С.Дж. (1991). Исследование logspline оценка плотности. Comput. Статист. Data Anal. 12 327–347.

Куперберг, К. и Стоун, С.Дж. (1992). Оценка плотности бревен создание цензурированных данных. J. Comput. График. Статист. 1 301– 328.

Маррон, Дж. С. и Рупперт, Д. (1994). Преобразования в ре- Из-за граничного смещения при оценке плотности ядра. Дж. Рой. Статист. Soc. Сер. B 56 653–671.

Маркс, Б. Д. и Эйлерс, PHC (1994). Прямое обобщенное объявление- Дитивное моделирование со штрафной вероятностью. Документ представлен на 9-м семинаре по статистическому моделированию, Эксетер, 1994.

Маркс, Б. Д. и Эйлерс, PHC (1996). Прямое обобщенное аддитивное моделирование со штрафной вероятностью. Не опубликовано рукопись.

МакКаллаг П. и Нелдер Дж. А. (1989). Обобщенный линейный Модели, 2-е изд. Чепмен и Холл, Лондон.

O'Салливан, Ф. (1986). Статистический взгляд на некорректно поставленный стихотворные задачи (с обсуждением). Статист. Sci. 1 505–527.

O'Салливан, Ф. (1988). Быстрое вычисление полностью автоматизированного журнала оценщики плотности и логарифма опасности. SIAM J. Sci. Статист. Comput. 9 363–379.

Райнш, К. (1967). Сглаживание сплайн-функциями. Нумер. Математика. 10 177–183.

Сакамото, Ю., Исигуро, М. и Китагава, Г. (1986). Акаике Статистика по информационным критериям. Рейдел, Дордрехт.

Скотт, DW (1992). Многомерная оценка плотности: теория, Практика и визуализация. Вили, Нью-Йорк.

Сильверман, Б.В. (1985). Некоторые аспекты гладкости сплайна подход к построению кривой непараметрической регрессии (с обсуждение). Дж. Рой. Статист. Soc. Сер. B 47 1–52.

Сильверман, Б.В. (1986). Оценка плотности для статистики и Анализ данных. Чепмен и Холл, Лондон.

Вахба, Г. (1990). Сплайновые модели для данных наблюдений. СИАМ, Филадельфия.

Палочка, член парламента и Джонс, MC (1993). Сглаживание ядра. Глава- Человек и Холл, Лондон.

Уиттакер, ET (1923). О новой методике градуировки. Proc. Edinburgh Math. Soc. 41 63–75.

Комментарий

ST. Чиу

Авторы Пол Эйлерс и Брайан Маркс предоставляют очень интересный подход к непараметрической кривой примерка. Они дают краткий, но очень лаконичный обзор

ST. Чиу из Департамента статистики,
Государственный университет Колорадо, Форт-Коллинз, Колорадо 80523-0001.

B-шлицы. Мне также понравилось читать ту часть, где авторы применили свою процедуру к некоторым примерам. Как показано в статье, подход имеет несколько достоинства, заслуживающие более подробного изучения.

Подобно любому непараметрическому сглаживателю, про- поставленной процедуре требуется параметр сглаживания λ, чтобы контролировать плавность аппроксимации кривой. Мой ком-

Стр. 15

вопросы в основном касаются выбора сглаживания параметр.

Как известно, классические селекторы, такие как поскольку AIC, GCV, Mallows's C_p и т. д. не дают удовлетворительный результат. Для случая регрессии больше подробности о дефектах можно найти в Rice (1984). и Чиу (1991a). Скотт и Террелл (1987) и Чиу (1991b) обсуждают случай оценки плотности. В классические селекторы имеют большую вариацию выборки и склонность к выбору небольшого параметра сглаживания, таким образом получается очень грубая оценка кривой. это естественно ожидать, что у них есть похожая проблема применительно к выбору параметра сглаживания для P-шлицев.

Было предложено несколько процедур для восстановления устраните недостатки классических процедур. Чиу (1996) дает обзор некоторых из этих новых селекторы для оценки плотности. К сожалению, В этом случае некоторые процедуры предлагаются в Чиу. (1991a), Холл и Джонстон (1992) и Холл, Мар- Рон и Парк (1992).

Ниже я сделаю краткий обзор бывших выявите дефекты и исправьте классические селекторы для оценки регрессии ядра. Давайте как- возьмем простейшую модель круглой конструкции с на равном расстоянии друг от друга. $y_t = \mu x_t + \varepsilon_t$, где ε_t - это iid шум. Для ядерной оценки $\hat{\mu}_\beta$ с ширину полосы β, мы часто используем среднее значение суммы квадратичные ошибки

$$R\beta = E\left[\sum \hat{\mu}_\beta x_t - \mu x_j{}^2\right] \qquad (1)$$

для измерения близости между $\hat{\mu}x$ и μx.

Цель выбора полосы пропускания - выбрать оптимальная ширина полосы, которая минимизирует $R\beta$. Поскольку в практика μ неизвестна, мы должны оценить $R\beta$ и используем минимизатор оцененного $R\lambda$ как оценка оптимальной пропускной способности. К экзамену- Ple, Мэллоуз C_p имеет вид

$$\hat{R}\beta = RSS\beta - T\sigma^2 + 2\sigma^2 w0 / \beta \qquad (2)$$

Здесь wx - ядро, а σ^2 - дисперсия ошибки. Другие классические процедуры, такие как AIC и GCV имеют аналогичный вид и показаны как асимптотические полностью эквивалентен у Райса (1984). Все эти про- cedures полагаются на остаточную сумму квадратов $RSS\beta$.

Mallows (1973) предложил процедуру, основанную на наблюдение, что

$$R\beta = ERSS\beta - T\sigma^2 + 2\sigma^2 w0 / \beta$$

Как мы объясним позже, основная проблема здесь в что $RSS\beta$ не является хорошей оценкой ожидаемого значение.

Используя преобразование Фурье, (1) и (2) могли бы записываться соответственно как

$$R\beta = 4\pi \sum_{j=1}^{N} I_S \lambda_j \quad 1 - W_\beta \lambda_j{}^2 + \sigma^2 \sum_{j=1}^{N} W_\beta \lambda^2 + \sigma^2 \qquad (3)$$

а также

$$\hat{R}\beta = 4\pi \sum_{j=1}^{N} \{I_Y \lambda_j - \frac{\sigma^2}{2\pi}\} \quad 1 - W_\beta \lambda_j{}^2 + \sigma^2 \sum_{j=1}^{N} W_\beta \lambda^2 + \sigma^2 \qquad (4)$$

где I_Y и I_S - периодограммы Y_t и сигнал $S_t = \mu t / T$ соответственно и $\lambda_j = 2\pi j / T$, j = 1 N = T / 2. Кроме того, $W_\beta \lambda$ - это передача функция wt / βT / βT.

Сравнивая (3) и (4), видим, что $\hat{R}$ пытается используйте $I_Y \lambda - \sigma^2 / 2\pi$, чтобы оценить $I_S \lambda$. Сложность состоит в том, что на высокой частоте в I_Y преобладает шум и , таким образом , не дает хорошую оценку I_S .

Чиу (1991a) предложил обрезать частотная часть при оценке $R\beta$,

$$\tilde{R}\beta = 4\pi \sum_{j=1}^{J} \{I_Y \lambda_j - \frac{\sigma^2}{2\pi}\} \quad 1 - W_\beta \lambda_j{}^2 + \sigma^2 \sum_{j=1}^{N} W_\beta \lambda^2 + \sigma^2 \qquad (5)$$

Здесь J выбрано таким образом, чтобы не было никакой сигнатуры. щественны I_S за частотой $Л_J$. Селектор $\tilde{R}\beta$ имеет гораздо лучшую производительность, чем классический единицы. Холл, Маррон и Парк (1992) предложили другая процедура, уменьшающая вклад ция от высокочастотной части.

Понятно, что основы ядерной регрессии являются синусоидальными волнами. Основная причина успеха критерием (5) является то, что большая часть информации о μ концентрируется на низкой частоте. Другими словами, мы просто нужно довольно много баз, чтобы приблизиться к истинному кривой хорошо.

Однако, поскольку каждый базис B-сплайна очень локально для определенного интервала, мы не можем использовать только несколько оснований для аппроксимации кривой по всей площади Гион. На мой взгляд, это может стать большим препятствием для понимание и улучшение классического селекторы параметров сглаживания.

РЕКОМЕНДАЦИИ

Чиу, С.-Т. (1991a). Некоторые селекторы стабилизированной полосы пропускания для не- параметрическая регрессия. Аня. Статист. 19 1528–1546.

Стр.16

Чиу, С.-Т. (1991b). Выбор полосы пропускания для оценки плотности ядра связь. Аня. Статист. 19 1883–1905 гг.

Чиу, С.-Т. (1996). Сравнительный обзор выбора полосы пропускания для оценки плотности ядра. Статист. Sinica 6 129–145.

Холл, П. и Джонстон, И. (1992). Эмпирические функционалы и эффективный выбор параметров сглаживания. Дж. Рой. Статист. Soc. Сер. B 54 519–521.

Холл, П., Маррон, Дж. С. и Парк, БЮ (1992). Сглаженный крест Проверка. Вероятно. Области, связанные с теорией 92 1–20.

Маллоуз, К. (1973). Некоторые комментарии к C_p . Технометрика 15 661–675.

Райс, Дж. (1984). Выбор полосы пропускания для непараметрической регрессии. Аня. Статист. 12 1215–1230.

Скотт, Д.В. и Террелл, Г.Р. (1987). Предвзятый и непредвзятый перекрестная проверка при оценке плотности. J. Amer. Статист. Доц. 82 1131–1146.

Комментарий

Дуглас Ничка и Дэвид Камминс

Одной из сильных сторон презентации авторов является простые формулы регрессии гребня, которые приводят к оценщик. Мы хотели бы указать на разложение с использованием другого набора базовых функций, которые помогает интерпретировать это более плавно. Эта альтернатива основание, полученное из B-сплайнов, облегчает вычисление оценка функции GCV и доверительных интервалов для расчетная кривая.

Чтобы упростить обсуждение, предположим, что $W = I$, поэтому что матрица шляпы

$$H = BB^T B + \lambda D^T D^{-1} B^T = GI + \lambda^{-1} G^T$$

$G = BQ_2^{-1/2} U$, $Q_2 = B^T B U$, $=$ diagγ и U равно ортогональная матрица такая, что $Q_2^{-1/2} D^T D Q_2^{-1/2}$ знак равно UU^T. Столбцы G можно отождествить с новый набор функций, известный как Demmler–Основа Рейнша (DR). В частности, это кусочно полиномиальные функции, ψ_v так что элементы G удовлетворяет $\psi_v x_i = G_{iv}$. Помимо полезного или-По свойствам тогональности базис ДР может быть заказан по частоте и при больших значениях γ_v будет наблюдаться больше колебаний (фактически v - 1 пересечений нуля). Рисунок-ure 1 (a) отображает несколько базисных функций для m = 133 равноотстоящих крестика и 20 равноотстоящих внутренние узлы. Рисунок 1 (b) иллюстрирует ожидаемый полиномиальное увеличение размера γ_v как функция из v.

Базис Деммлера – Райнша дает информацию информативная интерпретация сплайновой оценки. Пусть $\hat{f}$ de-

Дуглас Ничка - профессор статистики и Дэвид Cummins работает с Департаментом статистики, Север Государственный университет Каролины, Роли, Северная Каролина 27695-8203.

отметим P-сплайн, и пусть $\alpha = G^T y$ обозначает наименьшее квадраты коэффициентов от регрессии y на основе DR sis функции:

$$\hat{F}x_i = Hy_i = GI + \lambda^{-1} G^T y_i$$

$$\text{знак равно} \sum_{v=1}^{n} \psi_v x_i \frac{\alpha_v}{1 + \lambda\gamma_v}$$

Обратите внимание, что сглаживание - это просто линейная комбинация. наложение базисных функций DR с помощью коэффициентов которые уменьшаются (или сужаются) по факту тор $1/1 + \lambda\gamma_v$ из оценок наименьших квадратов. Из-за связи между γ_v и ψ_v (см. рисунок 1), базисные функции, представляющие высокочастотная структура будет иметь коэффициенты которые имеют более низкий вес. Этим способом более гладкий - это фильтр нижних частот, стремящийся сохранить низкочастотная структура и утяжеление высших частота терминов. Остаточная сумма квадратов и след H может быть вычислен быстро (порядок n) используя представление DR. Таким образом, функция GCV-также можно оценить в порядке n операций для заданное значение λ.

Еще одно применение формы DR - вычислительная техника. в полосе уверенности. Рассмотрим набор кандидатов функции, которые содержат истинную функцию с правильный уровень уверенности. Полоса уверенности тогда конверт подразумевается с учетом всех функций в этом наборе. Например, пусть $\hat{f}$ обозначает функцию и для $C_1 C_2 > 0$ положим

$$= \{h h \text{ - B-сплайн с коэффициентами } b, \sum_{я=1}^{n} \hat{F}x_i - hx_i{}^2 \leq C_1 \text{ и } b^T D^T D b \leq C_2\}$$

Рис. 1. Иллюстрация нескольких базисных функций Деммлера – Рейнша и связанных с ними собственных значений для 20 равноотстоящих узлов, 133 одинаково разнесенные наблюдения и вторые разделенные разности k = 2: верхний график (a) - ψ_v для ν = 3 5 10 15; цифры обозначают порядок этих базисных функций и на втором графике (b) идентифицируют собственные значения для этих функций.

Константы C_1 и C_2 определяются так, чтобы
Pf ∈ равняется желаемому уровню уверенности. В верхняя и нижняя границы доверительной полосы тогда

Ux = maxhx h ∈

а также

Lx = minhx h ∈

На практике мы работаем с коэффициентами и таким образом, вычисление U и L для каждого x равно задача минимизации с двумя квадратичными кон-

напряжения. Использование базиса DR снижает оба контекста. напряжения к квадратичным формам с диагональными матрицами и, таким образом, оба вычислимы в порядке n операций. ции. Причем эта стратегия не зависит от штраф за шероховатость разделяется на различия, но будет работать для любой неотрицательной матрицы, используемой в качестве штраф (например, шлицы тонкой пластины). В настоящее время мы исследуя выбор C_1 и C_2 на основе Оценка GCV f.

ПОДТВЕРЖДЕНИЕ

Работа поддержана NSF Grant DMS-92-. 17866.

Стр.18

Комментарий

Чонг Гу

Хочу начать с поздравления ау- Тора Эйлерса и Маркса за четкое изложение интересный вариант штрафных регрессионных сплайнов. Мои комментарии касаются трех вопросов: Р-шлицы действительно лучше? Что делает оптимальный гладкий означает? И что ждет в будущем непараметрическая оценка функции?

ДЕЙСТВИТЕЛЬНО ЛУЧШЕ Р-СПЛИНОВ?

Р-шлицы, безусловно, могут быть такими же полезными, как и другие варианты муравьи штрафных регрессионных сплайнов, но я не уверен, что они действительно выгодны по сравнению с другими эры. Это правда, что при огромных размерах выборки можно выберите n намного меньше m, чтобы сэкономить на вычислениях. без ущерба для производительности, но другие варианты муравьи регрессионных сплайнов также имеют одни и те же адреса - преимущество. Механическое обращение с разницей штраф, безусловно, очень интересное вычисление - союзник, но что касается конечных пользователей, я не понимаю, почему необходимы дискретные штрафы- значительно выгоднее непрерывных. Выше- штрафы за производные финансовые инструменты, безусловно, также возможны как дискретные штрафы в вычислительном отношении, хотя и более сложно реализовать, но разница несущественна - выгодно для конечных пользователей, чей основной интерес интерфейс.

Пользователи могут быть больше заинтересованы в том, что gram вычисляет, а не как вычисляет, как- штраф действительно действует, кроме того, что он повторно уменьшить эффективный размер в некоторых не так легко приемлемый способ.

Штрафные сглаживания с квадратичными штрафами: известно, что эквивалентно байесовским оценкам с Приоры по Гауссу. Когда $Q = D^T_k D_k$ полного ранга, соответствующий априор для коэффициентов B-сплайна а имеет среднее значение 0 и ковариацию, пропорциональную Q^{-1}. Когда Q имеет недостаточный ранг, у предшествующего есть «фиксированный эффect »рассеивается в нулевом пространстве Q и компонент «случайного эффекта» со средним 0 и ко- дисперсия, пропорциональная Q^+, показатель Мура – Пенроуза стих Q. С этой точки зрения P-сплайны различаются из других вариантов штрафных регрессионных сплайнов только в спецификации Q.

ЧТО ДАЕТ ОПТИМАЛЬНОЕ Сглаживание СТАТЬ ЗА?

Вероятно, важность никогда не переоценить выбора параметров сглаживания для любого успеха- Полное практическое применение любого метода сглаживания. AIC и перекрестная проверка являются одними из наиболее эффективных. принятые (и успешные) рабочие критерии для модели отбора, но их оптимальность установлена, теоретически или эмпирически, только для конкретной проблемы настройки при соответствующих условиях. Наивный адап- Использование этих критериев в новых постановках задач не обязательно обеспечивать почти оптимальную посадку.

когда-либо, и в этом отношении я только вижу, что Р-шлицы проигрывают до штрафных регрессионных сплайнов с обычным производные штрафы, понятные каждому.
Говорят, что B-шлицы обеспечивают хорошую основу для аппроксимации функции пользователи могут просто игнорировать любые другие свойства B-сплайнов и по-прежнему иметь четкое представление о том, что они собой представляют получение от производных штрафов или, для этого мат-тер, из дискретных штрафов Уиттекера, которые используют различия значений соседних функций. С участием Р-шлицы, однако, интуиция неудовлетворительна. в последнее время отняли у пользователей, и даже с доскональное знание всех свойств B-шлицы, я не уверен, что можно легко понять, что

Чонг Гу - доцент кафедры
Статистика, Университет Пердью, Вест-Лафайет, Инди-
Ана 47907.

В частности, меня несколько беспокоит «Оптимальность» наивных приспособлений этих критических teria, провозглашенная в разделе 6. Во-первых, в в каком смысле эти критерии являются «оптимальными» в данной проблеме? параметры lem, к которым они применяются; во-вторых, там не является эмпирическим (или теоретическим) свидетельством, иллюстрирующим предполагаемая «оптимальность». AIC или перекрестная проверка могут обеспечить почти оптимальную посадку, но они, безусловно, подходят сами по себе не определяют понятие оптимальности.

Мои опасения проистекают из предыдущих эмпирических исследований. экспериментов с выбором параметра сглаживания по я и другие, особенно в негауссовских проблемы регрессии (обычно называемые ген-обобщенные линейные модели). Использование Кульбака – Лейблера несоответствие или его симметризованная версия для определения оптимальности, было обнаружено, что наивная адаптивная оценка GCV в негауссовой регрессии, которая похоже на то, что предлагают авторы в разд. ция 7, может дать что угодно, кроме почти оптимальной посадки. См., Например, Cox and Chang (1990), Gu (1992).

Стр.19

и Сян и Вахба (1996). Для оценки плотности в разделе 8, я не смог найти определение матрицы H для понимания AIC предложено, но что бы это ни было, оно должно быть предметом к такому же вниманию, прежде чем быть рекомендованным как «Оптимально».

В обычной гауссовой регрессии оптимальность GCV хорошо известна в литературе. Для оценка AIC, представленная в (27), однако я бы как какое-то эмпирическое свидетельство, чтобы убедиться в его оптимальность. Скептицизм частично вызван некоторыми эмпирические данные, позволяющие предположить, что след H не может быть последовательной характеристикой эф-Фактический размер модели. Такие доказательства могут можно найти в Gu (1996), доступном в Интернете по адресу http: // www.stat.lsa.umich.edu/~chong/ps/modl.ps.

ЧТО ДЕНЬ ГРЯДУЩИЙ ДЛЯ ОЦЕНКИ ФУНКЦИЙ?

В ответ на желание статистической науки для размышлений о будущих направлениях исследований-я хотел бы воспользоваться этой возможностью, чтобы предложить некоторые из моих мыслей.

Давно сказано, что все методы сглаживания работают аналогично в одном измерении, при условии, что выбор параметра сглаживания осуществляется правильно. верно, но снова и снова новые и не такие новые методы-оды продолжают изобретать. Настоящая проблема, как-всегда, кажется, заключается в многомерных задачах. Среди проклятие размерности и потенциальных структур связанных с многомерными задачами, выбор методов могут существенно повлиять на значение, в простоте вычислений и сглаживания выбор параметра, для удобства использования проработка конструкций и т. д. Среди методов с наибольшим потенциалом являются адаптивные регрессии. зубчатые шлицы, разработанные Фридманом, Стоуном и соавторами. рабочие и сглаживающие шлицы, разработанные школа сплайнов в Висконсине, которую ведет Вахба. Пе-Тем не менее, сплайновый подход с окончательной регрессией кажется несколько затруднен отсутствием эффективных ба-sis, скажем, в двух или трех измерениях.

Еще более сложная задача - важная линия переоборудования. поиск, которым в основном пренебрегали, - это умозаключение. Что обычно получается из оценки функции в литературе представлены точечные оценки, возможно, с асимптотическими скорость сходимости и интуитивное сглаживание селекторы параметров не всегда сопровождаются обоснования. Помимо нескольких записей, основанных на Байесовская модель сглаживающих сплайнов Вахбы (1983), Кокс, Кох, Вахба и Янделл (1988), Барри (1993) и некоторые последующие действия, практические процедуры, которые оценки интервалов, проверка гипотез и т. д. по, в значительной степени отсутствуют в литературе. Охранять против опасности чрезмерной интерпретации данных использование непараметрических методов, таких как выводной инструменты должны стать главным приоритетом в будущих исследованиях. В модели Байеса, где целевая функция рассматривается как реализация случайного процесса, развитие может продолжаться в рамках обычного логическая основа. При традиционном наборе-ting, где целевая функция считается фиксированной, однако, возможно, придется отвернуться от контраргумента. традиционное мышление Неймана – Пирсона, прежде чем он сможет называйте любые полезные инструменты вывода не-ad-hoc.

РЕКОМЕНДАЦИИ

Барри, Д. (1993). Проверка аддитивности функции регрессии. Аня. Статист. 21 235–254.
Кокс, Д.Д. и Чанг, Ю.-Ф. (1990). Итерированное пространство состояний al-горитмы и перекрестная проверка для обобщенного сглаживания шлицы. Технический отчет 49, Департамент статистики, Univ. Иллинойс.
Кокс, Д.Д., Кох, Э., Вахба, Г. и Янделл, Б.С. (1988). Проверка (параметрической) гипотезы нулевой модели в (полупараметрической) метрические) частные и обобщенные сплайн-модели. Аня. Статист. 16 113–119.
Гу, С. (1992). Перекрестная проверка негауссовских данных. Журнал Вычислительная и графическая статистика 1 169–179.
Гу, С. (1996). Выбор параметров индексации и сглаживания модели в оценке непараметрических функций. Технический отчет 93-55 (ред.), Департамент статистики, Purdue Univ.
Вахба, Г. (1983). Байесовские «доверительные интервалы» для перекрестных проверенный сглаживающий сплайн. Дж. Рой. Статист. Soc. Сер. B 45 133–150.
Сян Д. и Вахба Г. (1996). Обобщенное приближенное перекрестная проверка для сглаживания сплайнов с негауссовыми Дата. Статист. Синица. Появляться.

Стр.20

Комментарий

МС Джонс

Эйлерс и Маркс представляют ясную и интересную объяснение их методологии сглаживания P-сплайна. Ясно, что P-сплайны представляют собой еще один респектабельный подход к сглаживанию. Однако их хорошая опора Похоже, что в целом наравне с различные другие подходы; метод не ближе до или дальше от «быть идеальным более гладким», чем другие.

«P-шлицы не имеют граничных эффектов, они прямое расширение (обобщенных) линейных пере- модели агрессии, сохраняющие моменты (значит, вари- ances) данных и имеют полиномиальную кривую, подходящую как пределы. " За исключением третьего пункта те же утверждения может быть выполнено сглаживание сплайнов (Green и Sil- verman, 1994) или локальной полиномиальной аппроксимации (Fan and Gijbels, 1996).

Сохранение моментов кажется неважным. В регресс, я не вижу желательности. По плотности оценка, простые поправки плотности ядра es- сроки для инфляции дисперсии существуют, но разница от нормальной плотности (Джонс, 1991). Действительно, правильное использование средств и отклонений концепция, основанная на нормальности, поэтому исправленное ядро оценки действуют в нормальном полупараметрическом манера. Эфрон и Тибширани (1996) предлагают больше сложное сохранение момента, но первоначальный инди- катионы в том, что это не лучше и не хуже, чем все внутренние полупараметрические оценки плотности (Hjort, 1996).

«Расчеты, в том числе для кросс- валидации, относительно недороги и легко включены в стандартное программное обеспечение ». Опять же, про- компоненты двух конкурирующих методов, которые у меня есть упомянутые будут утверждать то же самое в первой половине этого и сторонники регрессионных сплайнов будут требовать лот.

Особо нового вклада авторы не вносят. Возможность автоматического выбора полосы пропускания. Крест- проверка и AIC относятся к классу методов (например, Härdle, 1990, стр. 166–167), который, хотя и не был все очень плохо, позвольте возможности для улучшения.

М.К. Джонс - читатель по статистике, отделение Управление статистики, Открытый университет, Уолтон-холл, Милтон Кейнс, MK7 6AA, Великобритания.

Быстрый расчет этих селекторов полосы пропускания менее важно, чем разработка лучших селекторов. Для для локальных многочленов предлагаются улучшения (для нормальные ошибки) Fan and Gijbels (1995) и Rup- Перт, Шизер и Палочка (1995) и неопубликованные работа распространяет их на более общие ситуации.

Сравнение (5) с (11) сосредоточено на небольшая дополнительная сложность последнего. Но что это более интерпретируемый: штраф за шероховатость на кривой или по серии коэффициентов? Изменение штрафа в настройке сглаживающего сплайна позволяет различные пара- метрические пределы (например, Ansley, Kohn and Wong, 1993); как P-шлицы могут с этим справиться?

Раздражающий аспект сплайнового подхода - es - отсутствие прямого (асимптотического) среднего квадрат ошибки - тип результатов для обозначения теоретической производительность cal относительно ядра / локального полинома подходы, для которых такие результаты просто получаются содержательный и, в определенных пределах, информативный. Я сомневаюсь могут ли P-шлицы способствовать таким разработкам (причина указана ниже).

Кажется, что у P-сплайнов нет особой привлекательности. гибкость для многомерных приложений. Примеры заслуживают внимания только тем, что выглядят как полученные результаты - можно и другими способами.

Идея оценки плотности P-сплайнов трактовать точное бинирование как регрессию Пуассона данные. Хорошо, но опять же в равной степени применимо к другим подходов и уже исследованы для локальных поли- номинальное сглаживание. Симонов (1996, раздел 6.4) и Джонс (1996) объясняют, как такая регрессия подходы к оценке плотности дискретизированы версии некой «прямой» локальной плотности правдоподобия методы оценки (Hjort, Jones, 1996; Loader, 1996). Биннинг - основное вычислительное устройство все оценки типа ядра (Fan and Marron, 1994). Подход локального правдоподобия уже глубоко понял теоретически.

Сравнение разумной границы P-сплайнов производительность с разумными локальными многочленами Граничная производительность еще не доступна через теория или моделирование.

Интересный момент, упомянутый в статье, - кажущийся континуум между пара- метрика подходит с одной стороны и полностью «непараметрическая» с другой стороны, с многопараметрическими пара-

Стр.21

метрические модели и полупараметрические подходы в
между: дихотомия на параметрические и непараметрические
метрика неуместна, и есть огромная серая
площадь перекрытия. Эквивалентные степени свободы
идеи Хасти и Тибширани (1990)
(но возможно улучшить?) попытаться дать это согласие
тиниум шкала. Теоретическая разработка может быть
усложняется P-образными шлицами по причинам, связанным с
связан с количественной оценкой «непараметрическости»
промежуточные методы.

Наконец, мы вернемся к моей основной мысли. В рекламе
чудесный «личный взгляд на сглаживание и статистику»,
Маррон (1996) дает список методов сглаживания
и еще один из факторов (к которому я мог бы добавить другие)
участвует в выборе между методами. Маррон
говорит: «Все методы перечисленные иметь разные
сильные и слабые стороны в расходящиеся чувства.
Ни один из этих методов не доминирует над другими во всех отношениях.
чувств. Поскольку эти факторы настолько разные -
Ent, почти любой метод может быть «лучшим», просто
соответствующий личный вес различных фак-
задействованы враги ". P-шлицы - разумное дополнение
в первый список Маррона, но не имеют особого статуса
что касается его второго.

РЕКОМЕНДАЦИИ

Ансли, К.Ф., Кон, Р. и Вонг, К.М. (1993). Непараметрический сплайн-регрессия с априорной информацией. Биометрика 80 75–88.

Эфрон, Б. и Тибширани, Р. (1996). Используя специально разработанные экспоненциальные семейства для оценки плотности. Аня. Статист. 24 000–000.

Fan, J. и Gijbels, I. (1996). Локальное полиномиальное моделирование и Его приложения. Чепмен и Холл, Лондон.

Фан, Дж. И Маррон, Дж. С. (1994). Быстрое внедрение не-параметрические оценщики кривых. J. Comput. График. Статист. 3 35–56.

Hjort, NL (1996). Выступление полуфабрикатов Эфрона и Тибширани. параметрическая оценка плотности. Неопубликованная рукопись.

Хьорт, Н.Л. и Джонс, М.К. (1996). Локально параметрические не-параметрическая оценка плотности. Аня. Статист. 24 1619–1647.

Джонс, МС (1991). О поправке на инфляцию дисперсии в кер-оценка плотности nel. Comput. Статист. Data Anal. 11 3–15.

Джонс, МС (1996). О близких отношениях локальной плотности правдоподобия оценка. Неопубликованная рукопись.

Погрузчик, CR (1996). Оценка локальной плотности правдоподобия. Аня. Статист. 24 1602–1618

Маррон, JS (1996). Персональный взгляд на сглаживание и статистику тики (с обсуждением). Comput. Статист. Появляться.

Рупперт, Д., Шезер, С.Дж. и Ванд, член парламента (1995). Эф-эффективный селектор полосы пропускания для локальной регрессии методом наименьших квадратов. J. Amer. Статист. Доц. 90 1257–1270.

Симонов, JS (1996). Методы сглаживания в статистике. Спрингер, Нью-Йорк.

Комментарий

Иоахим Энгель и Алоис Кнайп

Пол Эйлерс и Брайан Маркс предоставили нам
с красивым и гибким дополнением к сглаживанию
Инструментарий. Предложенная ими оценка P-сплайна может быть
рассматривается как некий компромисс между обычными
Оценка B-сплайна и сглаживающий сплайн ар-
проехать. В отличие от многих работ по B-сплайнам,
однако они не рассматривают деликатную проблему
оптимального выбора узла. Вместо этого они предлагают
использовать большое количество равноотстоящих узлов. Сглаживание
вводится штрафом за шероховатость на разных
Сплайновые коэффициенты.

Оценка P-сплайна эквивалентна сглаживанию
шлицев при выборе как можно большего количества узлов

Иоахим Энгель работает с Wirtschaftstheorie II, Universität Bonn и математический факультет PH Людвигсбург, Германия. Алоис Кнайп с Institut de Statistique, Католический университет Лувена, Бельгия.

наблюдения n = m с узлом на каждом
точка данных. Однако это не та ситуация.
авторы имеют в виду. Предлагают выбрать
большое количество узлов n, но n <m. Такой ап-
Методика представляет значительный интерес. Мы знаем из
личный опыт, что непараметрическая регрессия
посадки на основе B-шлицев часто визуально более удобны.
звонка, чем, например, оценки ядра. В
то же самое, кажется, верно и для P-шлицев, если современный
используется количество узлов. Кроме того, как
Авторы указывают, что P-шлицы вместе с диф-
штрафные санкции пользуются многими важными практическими рекомендациями.
преимущества и достаточно гибкие, чтобы применяться в
различные ситуации моделирования, например, дополнительно
или самомоделирующей регрессии, где
используется алгоритм подгонки.

Тем не менее, мы пока не видим много доказательств
для утверждения авторов, что P-шлицы «приближаются
быть идеальным более гладким ». Например, локальные поли-
номинальная регрессия, как известно, не имеет границ

Стр. 22

проблемы (в первую очередь) и обладать определенными оп-
временность и минимаксные свойства (Fan, 1993). Для
оценка плотности показывают, что Энгель и Гассер (1995)
минимаксное свойство фиксированной полосы пропускания ядра
в большом классе оценщиков содержат:

могло бы быть совсем иначе, если бы n m. Действительно, оценка
помощник может решающим образом зависеть от n. Следовательно, почему
не определять λ и n перекрестной проверкой или повторной проверкой.
поздний метод? Следующие теоретические аргументы
может предположить, что такая процедура подействует. Примечание

использование оценщиков правдоподобия со штрафными санкциями. Представленные бумага не дает никаких аргументов, ни тео-
ретикальным или моделированием, поддерживая любые превосходные превосходство P-шлицев над своими многочисленными конкурентами.

В случае регрессии теоретические свойства P-шлицев можно оценить, комбинируя ar-предложения де Бура (1978) об асимптотическом смещении и дисперсия B-сплайнов в (в зависимости от m, порядок сплайна k и гладкость нижележащего функция) с известными результатами по сглаживанию шлицы.

Авторы предлагают использовать AIC или кросс-проверка для выбора параметра сглаживания λ. Однако внимательный взгляд на их метод показывает, что что на самом деле есть два свободных параметра: λ и количество узлов n. Если n ≈ m, то необходимо получить сглаживающий сплайн, в то время как результаты что AIC и перекрестная проверка очень близки к не-предвзятая оценка риска, которая состоит из оценки оптимальные значения λ и n путем минимизации

$$\sum_{я=1}^{m} \left(y_i - \hat{\mu}_i\right)^2 + 2\sigma^2 \operatorname{tr} H_{\lambda n}$$

где $H \equiv H_{\lambda n}$ - соответствующая более гладкая ма-трикс. Обозначим через ASEλ n средний квадрат ошибки. Величина аппроксимации, полученная с использованием некоторых параметров λ и н. При определенных технических условиях он затем Из результатов Кнейпа (1994) следует, что при $m \to \infty$

$$\mathrm{ASE}\,\hat{\lambda}\,\hat{m} / \mathrm{ASE}\lambda_{opt}\, m_{opt} \to_P 1$$

Здесь $\hat{\lambda}$ и $\hat{m}$ і параметры, оцениваемые не-смещенная оценка риска, в то время как λ_{opt} и m_{opt} представляют прислал оптимальный выбор параметров минимиз-ing ASE.

Комментарий

Чарльз Куперберг

Эйлерс и Маркс представляют интересный подход сплайновому моделированию. В то время как оценка функции на основе на сглаживающих шлицах часто дает разумное повторное В результате вычислительная нагрузка может быть очень большой. Однако если количество базисных функций ограничено, вычисления становятся намного проще, и когда узлы равномерно расположены, решение действительно становится довольно элегантным. Чтобы повысить авторитет утверждения о том, что P-сплайны близки к «идеальному плавнее », необходимо решить несколько вопросов:

1. При оценке плотности, когда диапазон данных является ( ) полезно, чтобы оценка плотности была положительный на (+), например, для ресэмплинга. Некоторые методы позволяют оценивать плотности на ограниченных или неограниченные интервалы. P-шлицы не кажутся обладают этим свойством: нижняя и верхняя границы имеют подлежит уточнению, и кажется, что нет естественного способ экстраполировать за эти границы. Здесь как-нибудь обойти это? Может ли бесконечность быть ограничением?
   Как указать границы? От пример самоубийства кажется, что это может повлиять значительно улучшить результаты.
2. Чтобы использовать P-шлицы, необходимо сделал. Сколько узлов нужно использовать? Это процедура нечувствительна к количеству сучков при условии, что их достаточно? Если да, то как много достаточно? Как устроена вычислительная нагрузка зависит от количества узлов?
   Какой порядок штрафов следует использовать? Ты рекомендовать изучить несколько возможных наказаний, поскольку в примере логистической регрессии, или у вас есть другая рекомендация, например, использование k = 3 для оценка плотности, чтобы предел вашей оценки помощник при $\lambda \to \infty$ является нормальной плотностью? Поскольку многие процедуры сглаживания и оценки плотности используются в качестве инструментов EDA, хорошие значения по умолчанию очень важны - пока.
3. Было бы интересно увидеть применение методологии P-сплайна для более сложных данные, такие как данные о доходах, описанные ниже, который включает тысячи случаев, узкая вершина и серьезный выброс.
   Как бы алгоритм P-сплайна, где узлы расположены эквидистантно, ведут себя при есть серьезные выбросы, которые будут доминировать расположение узлов? Можно ли по-узлов неэквидистантно, например, на основе по статистике заказов?
4. Есть ли теоретические результаты о большой выборке? Какое поведение P-шлицев?

Чарльз Куперберг - доцент кафедры Департамент статистики Вашингтонского университета, Сиэтл, Вашингтон 98195-0001.

Стр. 23

ПОЛИНОМИАЛЬНЫЕ ШЛИФЫ И БЕСПЛАТНО ОЦЕНКА ПЛОТНОСТИ

Помимо подхода штрафной вероятности, существует это совершенно другой подход к функции es-(1996) дают обзор полиномиальных сплайнов. и их приложения.

Оценка плотности логарифмических линий, в которой (одномерный ate) логарифм-плотность моделируется кубическим сплайном, отображается обвинен в Куперберге и Стоуне (1992) и Стоуне и другие. (1996). Программное обеспечение для версии 1992 г., написанное на языке C и взаимодействует с S-PLUS, является общедоступным в состоянии из Statlib. (Версия LOGSPLINE 1992 г. использует только удаление узлов; здесь, однако, мы сосредотачиваемся на версии 1996 года, в которой используются оба узла и удаление узла.) LOGSPLINE может предоставить оценку сопрягается как на конечных, так и на бесконечных интервалах, и может обрабатывать данные, подвергнутые цензуре.

Итоги LOGSPLINE на Old Faithful данные и данные о самоубийствах очень похожи на данные о кор-ответ на результаты P-сплайнов [данные о самоубийствах пример у Куперберга и Стоуна (1992)]. Здесь расчет времени на шлицах. Тогда как для P-шлицев как количество, так и расположение узлов фиксируются заранее, а плавность регулируется. дополненный параметром сглаживания, в полиноме шлицевой каркас количество и расположение узлы определяются адаптивно с использованием пошаговой мудрый алгоритм и без параметра сглаживания нужный. Такие методы полиномиальных сплайнов были используется для регрессии (Friedman, 1991), плотность времени (Куперберг, Стоун, 1992), полихото- мышечная (множественная логистическая) регрессия (Куперберг, Bose and Stone, 1997), анализ выживаемости (Kooper- berg, Stone and Truong, 1995a) и спектральной плотности. оценка города (Куперберг, Стоун и Чыонг, 1995b).

В методологиях одномерного полиномиального сплайна алгоритм начинается с довольно небольшого количества узлы. Затем он добавляет узлы в тех регионах, где дополнительный узел будет иметь наибольшее влияние, мы- статистика ing Rao (оценка), чтобы выбрать лучший место расположения; после заранее указанного максимального количества узлы достигнуты, узлы удаляются по одному, используя статистику Вальда, чтобы решить, какой узел повторно двигаться. Из последовательности подогнанных моделей одна с наименьшим значением критерия BIC выбрано.

Алгоритмы полиномиальных сплайнов для многомерных оценки функций аналогичны, за исключением того, что на каждом шаг сложения алгоритм добавляет узел в одна переменная или тензорное произведение двух или более одномерные базисные функции. Мы успешно при- применили такие методологии к наборам данных размером всего 50 для одномерной оценки плотности и столь же больших как 112,000 для 63-мерного полихотомического задача регрессии с 46 классами. Для неадап- Теоретические методологии использования полиномиальных сплайнов результаты относительно L 2 -скорости сходимости табличные. Стоун, Хансен, Куперберг и Чыонг

мы рассматриваем гораздо более сложный набор данных. В сплошной линией на рисунке 1 показаны значения плотности каротажных линий. время основано на случайной выборке из 7 125 годовых чистый доход в Соединенном Королевстве [Family Expen- diture Survey (1968–1983)]. (Данные были масштабируется так, чтобы получить среднее значение 1.) Девять узлов, которые были отмечены LOGSPLINE. Обратите внимание, что четыре из этих узлов находятся очень близко к пику около 024. Этот пик связан с пенсией по старости в Великобритании. Это привело к тому, что многие люди почти идентифицировали себя. реальные доходы. В «Куперберге и Стоуне» (1992) мы пришел к выводу, что высота и расположение этого пика точно оцениваются LOGSPLINE. Есть несколько причин, по которым эти данные более сложны чем старые верные и данные о самоубийствах: данные набор намного больше, так что это сложнее вычислительным ресурсам (оценка LOGSPLINE заняло 9 секунд на рабочей станции Sparc 10); ширина пика составляет около 0,02 по сравнению с диапазоном 11,5 данных; есть серьезный выброс (самый большой обслуживание - 11,5, второе по величине - 7,8); и рост плотности слева от пика очень отвесный.

Чтобы получить представление о том, как работают P-шлицы, dure для этих данных, я сначала удалил самое большое наблюдение, чтобы не было длинные пробелы в данных, уменьшая максимальное обслуживание до 7.8. Пунктирная линия на рисунке 1 - это Оценка LOGSPLINE по данным с фиксированными узлами при i / 20 × 78, для i = 0 1 20 (используя 20 меж- vals, как и в большинстве примеров P-сплайнов.) Подгонка должна быть аналогична подгонке P-сплайна с $\lambda = 0$. В этой оценке оказывается, что узкий пик полностью упущено, и это из-за крутого рост плотности слева от пика и отсутствие достаточно большого количества узлов у пика, два режимы оцениваются там, где существует только один режим.

Стр. 24

Рис. 1. Оценка плотности линий для данных о доходах (сплошная линия); x обозначают расположение узлов; логарифмическая аппроксимация оценка P-сплайна с параметром штрафа 0 (пунктирная линия).

Было бы очень интересно посмотреть, как Методология P-сплайна действует на этих данных, а в в частности, может ли он точно представлять острый пик около 0,24.

ПОДТВЕРЖДЕНИЕ

Исследование частично поддержано NSF Grant DMS- 94-03371.

Комментарий

Деннис Д. Кокс

Основная новая идея в этой статье - грубость штраф на основе коэффициентов B-сплайна. Там будут критики - я приведу некоторые критические замечания ниже - но в простоте идея. Если бы мне пришлось разрабатывать программное обеспечение ab initio, ясно, что предложенные штрафы за шероховатость здесь потребуется меньше усилий для реализации, чем стандартные, основанные на L_2 -норме секунды производная.

Прецедент использования B-сплайна коэффициентов таким прямым способом, с компьютера графика (CG) и компьютерное проектирование (CAD). «Контрольная точка» обычно используется в параметрических B-сплайн-представления кривых и поверхностей на основе Фактически состоит из коэффициентов B-сплайна. Увидеть Фоли и ван Дам (1995, раздел 11.2.3). Это демон-показано на рисунке 1, где контрольные точки для сплошная кривая - это просто случайная единообразная добавка к линейному ухо, и те же самые точки сужаются в сторону 0,5 перед добавлением тренда для получения контроля точки для пунктирной кривой. Ордината каждого контрольной точкой является кубический кардинальный коэффициент B-сплайна. cient, а абсцисса - середина опоры. В приложениях CG / CAD контрольные точки создаются. нипулированный, чтобы получить кривую или поверхность с желаемым форма или гладкость. Практикующие CG / CAD должны быть-ознакомьтесь с этими контрольными точками и разработайте

Деннис Д. Кокс из Департамента статистики, Райс Университет, почтовый ящик 1892, Хьюстон, Техас 77251.

Стр.25

Рис. 1. Пример контрольных точек: сплошная кривая происходит от сплошных контрольных точек, а пунктирная кривая - от треугольной контрольной точки. точки.

ощущение их влияния на кривую или поверхность. Точно так же статистики после некоторых усилий могут обнаружить что коэффициенты B-сплайна очень естественны.

Если бы у меня было такое же удобное программное обеспечение для сглаживания шлицы или P-шлицы, я бы предпочел первые, частично в основном из байесовских соображений. Байесовский интерпретация P-сплайнов (т. е. разностных B-Коэффициенты сплайна представляют собой гауссовский белый шум, не-der the Prior) более искусственна, чем обычная или как у Wahba (1978). В частности, обычный априорные значения указываются независимо от размера выборки, тогда как можно было бы использовать больше B-сплайнов с

Дискретная форма вариационного вывода в Спекман (1983) приводит к системе

$$\lambda_2 a_3 + \sum_i B_1 x_я \sum_j a_{кj} B_j x_i$$

$$= \sum_я y_я Б_1 x_я$$

$$\lambda_3 a_4 - \lambda_2 a_3 + \sum_i B_2 x_я \sum_j a_{кj} B_j x_i$$

$$= \sum_я y_я Б_2 x_я$$

больший образец. Кроме того, интеграл от вторую производную в квадрате легче интерпретировать из небайесовская перспектива, чем сумма квадратов вторых разностей коэффициентов B-сплайна.

Я не согласен с утверждением авторов о том, что их метод не имеет краевых задач. P-шлицы примерно эквивалентны сглаживающим сплайнам которые действительно имеют граничные эффекты (Speckman, 1983). Чтобы объяснить, рассмотрите возможность минимизации из уравнения (5),

$$Sa = \sum_{я=1}^{n} \{y_я - \sum_{j=1}^{n} a_j B_j x_i\}^2 + \lambda \sum_{j=3}^{n} {}^{2J} \quad {}^{2}$$

$$\lambda_4 a_к + \sum_я \quad Б_к x_я \sum_j \quad a_{к\,j} B_j x_i$$

$$= \sum_я \; y_я Б_к x_я \qquad 3 \le к \le п - 2$$

$$-\lambda_3 a_п - \lambda_2 a_п + \sum_i \qquad B_{п-1} \quad x_я \sum_j \quad a_{к\,j} B_j x_i$$

$$= \sum_я \; y_я Б_{п-1} \quad x_я$$

$$\lambda_2 a_п + \sum_i \quad Б_н x_я \sum_j \quad a_{к\,j} B_j x_i$$

$$= \sum_я \; y_я Б_н x_я$$

Стр. 26

Обратите внимание, что уравнения для коэффициентов около end включают разность более низкого порядка, поэтому меньше плавность наложена.

ПОДТВЕРЖДЕНИЕ

Исследование поддержано NSF Grant DMS-90-01726.

Комментарий

Стефан Р. Сайн и Дэвид В. Скотт

Нас интересовали формулировки проблема сглаживания, которые одновременно являются глобальными в природе с локально адаптивным поведением. Грубый-штрафные санкции, основанные на таких функционалах, как теграл квадратов вторых производных подобранного кривой пользовались большой популярностью. Решение к таким задачам оптимизации часто идет сплайн. В авторов следует поздравить с введением идея штрафовать за гладкость сплайна коэффициентов, что снижает размерность проблема, а также снижение сложности расчеты. В пользу этого подхода можно многое сказать.

Вообще интересно попробовать потренироваться формулировка эквивалентного ядра всех гладких методы. Это было сделано для Надарьи-Ватсон. сглаживание регрессии по Сильверману (1984), который продемонстрировал асимптотический способ, которым Оценщик адаптирован локально.

В настройке оценки плотности мы были изучение природы лучших локально адаптивных оценка плотности по линии Бреймана–Оценка Мейзеля – Перселла (Бреймана, Мейзеля и Пурселла). ячейка, 1977 г.)

$$(1) \quad \hat{F}x = \frac{1}{п} \sum_{я=1}^{n} \frac{1}{h_i} K\left(\frac{x - x_я}{привет_я}\right) = 1n_п \sum_{я=1} K_{h_i} \; x - x_я$$

Цель - найти оптимальный набор полос пропускания $\hat{H}_i$ i = 1 n, без ограничений на функцию

циональная форма. Сайн и Скотт (1996) исследуют при-proach, используя упакованную версию (1), где диапазон-ширина была найдена численно путем оптимизации более вариант метода наименьших квадратов или несмещенного перекрестного критерий валидации (UCV).

Удивительный результат нашего исследования заключается в том, что оптимальная оценка содержит явно нелокальную, поскольку

а также местные адаптивные функции. То есть полоса-ширины некоторых точек данных, особенно в хвосты очень большие. Это было довольно неожиданно поскольку Террелл и Скотт (1992) обсуждали негативные основные последствия такой большой пропускной способности, когда $h_i = h / \sqrt{fx_i}$, идея, предложенная Абрамсоном (1982) и широко изучены в литературе. Кроме того, Сайн и Скотт (1996) показали, что этот «закон квадратного корня» на практике лишен гибкости из-за зависимости исключительно от уровня не-плотность посадки. Отсылаем заинтересованного читателя к те статьи.

На рисунке 1 мы показываем три плотности гейзера. данные: (1) наша оптимальная локально адаптивная оценка; (2) а фиксированная оценка ядра (пропускная способность также выбирается UCV); и (3) авторский P-сплайн. Фиксированный подход к полосе пропускания не может найти ни одной полосы ширину, чтобы сгладить оба режима соответствующим образом, оставив правый режим недосглаженный. Более гибкий адаптивная оценка распознает локальную структуру нижележащей плотности и дает четкое представление отправка двух режимов в данных (отклонение возможность третьего режима) без лишних шум. Оценка P-сплайна дает оценку находясь где-то посередине между двумя подходами. это

Стефан Р. Саин работает в Южном методистском университете.

разнообразие. Дэвид В. Скотт, профессор статистики, Райс Университет, Хьюстон, Техас, 77251-1892.

Рис. 1. Расчетные плотности с использованием данных гейзера Old Faithful.

Стр. 27

интересно отметить, что адаптивный подход не чрезмерно сглажены, поскольку высоты двух режимов не затронут.

Р-сплайн (и другие подобные оценки) позволяют некоторая локальная адаптивность через механизм штрафа низм и ограничения на шероховатость посадки Кривая Теда. Однако выбор узлов (в данном случае сколь угодно большое количество равномерно расположенных узлов) может также повлиять на локальный характер оценки Р-сплайна. матор. Было бы интересно узнать больше о том, как выбрать узлы, в том числе «адаптивный» подход это может привести к более экономной модели, поскольку а также лучшее местное поведение. Наконец, надо выиграть- как выбрать ширину бункера для начального гистограмма влияет на итоговую оценку Р-сплайна. По нашему опыту, использование таких экспериментальных оценщиков может привести к плохим результатам, а также к затруднениям в самообслуживании. автоматическая реализация.

Нам было бы очень интересно увидеть более детальную Достаточно изучить, как Р-сплайны ведут себя по отношению к некоторым понятие оптимальной локальной адаптивности и как штраф и критерий AIC, а также другие параметры ters, можно настроить для достижения такого поведения.

ПОДТВЕРЖДЕНИЕ

Это исследование (второй автор) поддержано в часть предоставлена Грантом DMS-96-26187.

Возражать

Пол Х.К. Эйлерс и Брайан Д. Маркс

ВСТУПЛЕНИЕ

Заслуживают ли Р-шлицы места в центре внимания? Мы утверждал, что это вызвало много дискуссий. Мы благодарен за многие внимательные, позитивные и подробные Комментарии. Для нашего возражения мы сгруппируем их следующим образом; минимумы:

- дополнения и пояснения, особенно касающиеся обеспечение оптимального разглаживания;
- проблемы с производительностью Р-шлицев;
- сомнения по поводу нашего утверждения о том, что Р-шлицы приходят рядом с идеальной плавнее.

Мы будем реагировать в том же порядке, сначала комментируя расширения. Затем мы покажем анализы и экс- достаточно, чтобы показать, что мы можем решить все проблемы с стандартные Р-шлицы, кроме адаптивной гибкости (но необходимость в этом меньше, чем можно было бы подумать). Аф- что мы представляем своего рода «потребительский тест», таблица оценок для сравнения Р-образных шлицев с конкурирующими tition. Наконец, мы сделаем вывод, что Р-сплайны пересекаются почти все проблемы и резюмируйте, почему они привлекательный в использовании.

ОПТИМАЛЬНОЕ Сглаживание

Мы не имели в виду, что AIC и перекрестные валидация - это последнее слово об оптимальном плавном ing. Мы выступали за их использование, потому что они могут быть вычисляется легко и быстро, и поскольку у нас хороший опыт во многих реальных приложениях. Но поиск оптимальных критериев должен продолжаться, и в Р-шлицах нет препятствий для предотвращения использования более сложных методов.

Подход Чиу к фильтрации интересен. Тем не мение, кажется, что он ограничен равноудаленно выбранными данными, кроме потому что требуется (быстрое) преобразование Фурье.

Нычка и Камминс представляют интересный интерпретация сглаживания Р-сплайна как проекции на основе Деммлера – Райнша. Они используют равномерный неявно шагал по х-вектору. Для разреженных дизайнов некоторые столбцов В могут стать пустыми, в результате чего В т В в единственном числе. Мы подозреваем, что дополнительное ограничение - (например, небольшой штраф за гребень), тогда потребуется чтобы сделать возможным построение G. Преимущество- Базис Деммлера – Рейнша в основном со- ceptual: вычисление следа шляпы trix и GCV уже можно эффективно проводить с П-образные шлицы.

Пока мы обсуждаем эту тему, мы хотели бы добавим, что мы не понимаем широко распространенных озабоченность функционалом суммы квадратов ∫ ф - Ғ 2 dx как мера производительности в ден- оценка города и подробный анализ, было сделано. Можно было бы ожидать некоторого отклонения от нормы функциональное, или расстояние Кульбака – Лейблера, например Гу пользуется. В конце концов, никто не подходит по плотности к гистограмма с наименьшими квадратами. Кажется, что математика- Эматическая податливость - движущая сила этого, напоминая нам о пьянице, ищущей под уличный фонарь для ключей, которые он потерял где-то в темный. Чтобы помочь ему, умные математики подключают металлоискатель к фонарю.

Стр.28

Отметим, что для так называемых критических teria (Jones, Marron and Sheather, 1996), P-шлицы может быть очень полезным, когда нужно оценить (в-тегралы квадратов) третьих или более высоких производных, из-за простоты получения высокой степени B-шлицы. Но мы очень довольны AIC. Фигура 1 показаны данные Кука и Вайсберга (1994), что дает безжировая масса тела австралийских спортсменов. Эти данные были использованы в качестве испытательного стенда Джонс, Маррон и Шизер (1996). Мы используем AIC и получаем существенно такое же сглаживание, «второе поколение - результат по цене первого поколения ». Опять же, мы не хотят подразумевать, что AIC является окончательным ответом, но покажите, что это полезнее, чем иногда предложенный.

УЗЛЫ

В нашей статье мы были довольно консервативны в количество узлов, которые мы использовали и рекомендовали использовать. Еще возможно множество вариаций. Ниже мы увидим экс-довольно много сучков, даже больше, чем там точки данных, являющиеся контрпримером Энгелю и предположение Кнайпа. Они правы в том, что с столько узлов, сколько точек данных, мы приходим очень близко к сглаживающему сплайну, если x равноудаленный. Если это не так, нам понадобятся узлы на неэквидистантная сетка. Но тогда штраф должен тоже изменится: разделенные различия, например, $_{j}$ −a$_{j-1}$ / t$_{j}$ - т$_{j-1}$ в случае штрафа первой очереди должны быть использовал. Мы еще не полностью проанализировали эту ситуацию, но мы подозреваем некоторые интересные результаты, потому что B-сплайны на произвольной сетке вычисляются с схема разделенных разностей.

Рис. 1. Гистограмма и оценка плотности безжировой массы тела Австралийские спортсмены мужчины и женщины); 20 P-шлицев степени 3 штраф порядка 3 λ = 10.

Количество узлов в основном несущественно, так как пока он достаточно большой. Однако предложение Энгеля и Кнайпа для оптимизации λ и n могут иметь значение при стремлении к экономии. При λ = 0 мы можем измените n, пока не обнаружите, что для оптимального сглаживания появляется число от определенного n - 1 до n нужный. Возьмем n = n и увеличим λ для последней части. пути к оптимальности. Однако проблемы могут возникают с разреженными дизайнами, и в этом случае неидентифицируемые верность может наступить без штрафа.

ГРАНИЦЫ

Что касается заботы Куперберга о выборе границ, мы должны отличать «физическое» от «технологического» «конические» границы. Физическая граница определяется определяется природой исследуемой переменной. В В примере с самоубийством ноль - это физическая левая граница - да, потому что временные интервалы не могут иметь отрицательных длины, и поэтому не может быть плотности ниже нуля. Если мы выберем в качестве технической границы некоторую отрицательную значение, мы говорим, что могут быть отрицательные данные, что мы их не наблюдали, но это разумно. Можно оценить плотность в этом регионе. в В том же примере верхняя граница является технической, потому что мы не знаем верхнего предела (ну, может быть, 80 лет или около того). Неважно, какое значение берем, практический выбор в 2 раза высший наблюдаемое значение.

Куперберг спрашивает, как мы экстраполируем. Мы не делаем это: мы выбираем границы (когда физически значимый) достаточно широким, чтобы включать домен где требуется «экстраполяция».

Кокс показывает, что действительно существует граничный эффект, но более тонкого вида, чем мы думали в нашей газете, где мы имели в виду неприятный реквизит. количество ядер для распределения вероятностной массы вне (физические) границы в оценке плотности, или стремится к нулю в регрессии ядра.

ЗНАЧЕНИЯ ПАРАМЕТРОВ ПО УМОЛЧАНИЮ

Как справедливо замечает Куперберг, мы были расплывчаты в давая значения по умолчанию для некоторых параметров. Правила большой палец может быть следующим: возьмите интервал до-промежуточные узлы равны половине ширины самого узкого пика это должно быть видно на подобранной кривой; использовать B-шлицы 3-й степени и штраф 3-й степени. Рекомендуем графики AIC или GCV от эффективного размера, для порядки 1, 2 и 3 штрафа за разницу. Для оценки плотности путем сглаживания гистограмм мы рекомендуем поменять 100 или более ящиков. Это не может повредить большое количество B-шлицев, например 50 или 100, потому что штраф позаботится о любом переобучении. Из конечно, для более быстрых вычислений лучше иметь небольшое количество B-шлицев.

Стр.29

В некоторых случаях B-сплайны нулевой степени, которые просто постоянна между двумя узлами и ноль в другом месте, достаточно. Если принять количество узлов равным к количеству (эквидистантно выбранных) данных, B - единичная матрица. Мы не можем устоять перед темпом Чтобы показать, насколько простым становится сглаживание, способ. Пусть вектор y будет данными, а w вектором 0–1 весов для обозначения отсутствующих данных. Тогда три строки MATLAB

```
I = глаз (длина (y));
D = diff (I, 3);
```

сделать вывод, что большое количество наблюдений и выбросы не наносят ущерба P-образным шлицам, но что их фиксированная гибкость приводит к небольшому количеству сглаживания.

Доходы положительные и показывают большое соотношение между между максимумом и минимумом. Данные этого типа всегда следует изучать также в логарифмическом масштабе. Это показано на рисунке 4. Расчетная плотность выглядит очень разумно. Перегиб между двумя пиков указывается, что LOGSPLINE не выбирает вверх.

```
mu = (diag (w) + лямбда * D '* D) \ (diag (w) * y);
```

сделать трюк. На самом деле это просто Уиттакер (1923).
«Градационный» алгоритм.

АСИМПТОТИКА

У нас еще не было никаких асимптотических результатов по скорости сходимости. Однако заимствование из асимптотическая теория GLM, при сходимости при фиксированном λ асимптотическая дисперсия – ковариация матрица коэффициентов P-сплайнов имеет вид $\tilde{} = Q_B + Q_{\lambda}^{-1} Q_B Q_B + Q_{\lambda}^{-1}$. Этот результат особенно очень полезен для простого построения удвоенные полосы стандартных ошибок для $g\,\tilde{\mu}$, т. е. Варги $\tilde{\mu} = B \setminus B^{T}$. Другая асимптотическая теория минимумы относительно дисперсии и смещения P-сплайна коэффициенты. Конечно, их нужно переводить к свойствам расчетной кривой, потому что сами коэффициенты имеют ограниченную интерпретацию.

Рис. 2. Ширина ячейки гистограммы 01 и минимальное соответствие AIC 153 P-сплайны к данным о доходах; $\lambda = 10^{-6}$.

АДАПТИВНОЕ Сглаживание

P-шлицы обладают постоянной гибкостью. Куперберг, Сайн и Скотт обсуждают адаптивную оценку в какая непостоянная гибкость необходима. Сначала мы проанализируем данные о доходах, затем Old Faithful.

Мы должны признать с самого начала, что P-шлицы в их нынешнем виде не могут бросить вызов чрезвычайный контроль над гибкостью, которую LOGSPLINE из-fers. Тем не менее, мы можем пройти долгий путь с постоянной гибкостью. способность. На рисунке 2 показана гистограмма данных с ширина бункера 0,1 (из-за большого количества бункеров, счетчики отображаются в виде вертикальных линий посередине точку каждого бункера). Такой же интервал используется для узлы B-шлицев степени 3, дающие 153 их. У вас не может быть такого количества без штраф из-за серьезных проблем с идентифицируемостью. Мы видим, что оптимальная подгонка очень близка к его-тограммирует себя, слишком изгибаясь правым хвостом. Без выброса, мы получаем почти такой же результат, который следует ожидать с небольшим количеством гладкой ing. Рисунок 3 основан на части данных, используемых: бункеры меньшего размера. Левый пик скорее восстановлен. хорошо, но опять же правая часть кажется слишком неровной. Мы

Рис. 3. Ширина ячейки гистограммы 01 данных о доходах с $x < 2$ и минимальное соответствие AIC 53 P-сплайнов данным о доходах; $\lambda = 10^{-6}$.

Рис. 4. Ширина ячейки гистограммы 01 и минимальное соответствие AIC 53 P-сплайны к логарифму данных о доходах; $\lambda = 10^{-6}$.

Логарифмы также подходят для старых верующих. данные, как показано на рисунке 5. Мы оценили плотность для логарифмов данных и преобразовал, что вернуться к линейному масштабу. Обратите внимание, что хвост на

Конечно, будут ситуации, в которых простое преобразование не выйдет. Несговорчивый постоянная гибкость может быть реализована за счет использования подходящих веса в штрафе, например $\sum v_j \, d a_j^{2}$. Сим-Можно позаимствовать любой способ адаптивной оценки от Fan и Gijbels (1995): разделите поддержку на количество (перекрывающихся) интервалов, выполните плавное для каждого из них отдельно, давая номер оптимальных λ. Их можно интерполировать, чтобы получить гладкая кривая λ. Буквы v во взвешенном наказании можно установить пропорционально значениям этой кривой на узлах.

Fan et al. (1996) изучали ядерное сглаживание с помощью непрерывно изменяющаяся ширина полосы bx плавным интерполяция низкоразмерного набора точек $x_j b_j$, и оптимизируя b_j's. Можно представить аналогичным образом оптимизируя элементы v.

ПРОИЗВОДНЫЕ ИЛИ РАЗЛИЧИЯ?

Некоторые участники дискуссии предполагают, что штраф на

правая сторона намного короче, чем у Саина и
Скотт присутствует в их фигуре; это инфляция дисперсии
по ядру плавнее? Саин и Скотт упоминают
что они обнаружили неожиданно большую полосу ядра
ширины в хвостах. Левый хвост этой кривой
довольно странно, будучи заметно больше нуля
на большом протяжении.

Интересно отметить, что эффективный размер
Вычисленное нами значение составляет 10,8 для линейных данных и 8,2
для логарифмических данных, указывая, что на широте
Допускается значительно более сильное сглаживание шкалы.

вторая производная более ясна, чем одна по коэффициенту
знатоки B-шлицев. Мы думаем, что нет. Эти коэффициенты
высоты B-шлицев, которые создают посадку
кривая Теда, и поэтому у них есть прямой интуитивно понятный интерфейс.
претензия. Плавность требует, чтобы высота
соседние B-шлицы могут не сильно отличаться. В
штраф позволяет B-шлицам «держаться за руки», чтобы выдержать
беспорядочные колебания данных.

На рисунке 6 показаны смоделированные данные и соответствие B-сплайну.
без штрафа, в то время как на рисунке 7 сильное наказание
($\lambda = 2$, с разностями второго порядка). Примечание
гладкий конверт, который предлагают топы
B-шлицев.

Рис. 5. Гистограмма и минимальное соответствие AIC данных Old Faithful,
как в линейной, так и в логарифмической шкале с обратным преобразованием; в
Кривая с двумя режимами справа основана на линейном масштабе.

Рис. 6. Смоделированные данные, отдельные B-сплайны и подобранная кривая.
толстая линия без штрафа.

Стр.31

Рис. 7. Смоделированные данные, отдельные B-сплайны и подобранная кривая.
жирная линия со штрафом второго порядка и $\lambda = 2$.

Рис. 8. Скорости точек переменной звезды и фит 50 P-сплайнов.
со специальным штрафом, который заставляет подгонку приближаться к синусоиде.

Параметрический предел можно проиллюстрировать на
так же: когда вершины B-шлицев находятся на
прямая линия (парабола), аппроксимированная кривая
ухо (квадратное).

Как справедливо полагает Кокс, исходя из опыта САПР, с
немного практики можно развить хорошее интуитивное не-
понимание B-шлицев.

Мы согласны с Гу, что (очень либерально сказано)
«Наказание является предварительным». Мы также должны признать, что
там связь с деривативами намного яснее
чем один к разности высот B-шлица. В
в настоящее время мы можем указать только на близкую эквивалентность
обоих критериев, как показано в статье, но мы будем
ищите более интуитивное понимание.

наложить штраф, это тоже сработает, но
быть очень большими и очень плавными, качели вверх и вниз
в регионах без данных, потому что там только
гладкость, которая имеет значение. Это можно исправить
добавление небольшого штрафа за гребень, еще один пример
специализируясь на наказании за проблему.

Для полноты картины отметим, что Эйлерс (1991а)
использовали смешанные штрафы в задаче регрессии с
заказанные регрессоры. Эйлерс (1988) предложил использовать
штрафные наименьшие квадраты для оценки авторегрессии
модели для зашумленных сигналов с возможно отсутствующими данными.

СОХРАНЕНИЕ МГНОВЕНИЙ

Джонс не видит необходимости в сохранении
моменты регресса. Мы считаем важным:

СПЕЦИАЛИЗИРОВАННЫЕ ШТРАФЫ

Теперь мы подошли к штрафам за специальный параметр - ричными ограничениями и приведем пример, чтобы выразить нашу точку зрения. Предположим, что мы меняем штраф $\lambda \sum (a_j - 2a_{j-1} + a_{j-2})^2$ к $\lambda \sum (a_j - 2ca_{j-1} + a_{j-2})^2$, с $c = \cos 2\pi t / p$, а t - расстояние между узлами, то при высоких λ ряд a стремится к синусоидальной функции с периодом p: $a_j = a_0 \cos 2\pi jt / p + \varphi$, с a_0 и φ определяется по данным. Это заставляет подогнанный стремится к синусоидальному сигналу, интерполированному B-сплайнами. Если t мало по сравнению с p, это будет эффективно синусоида. На рисунке 8 показана часть серии измерения скоростей переменной звезды, см предполагалось иметь нулевое среднее значение; данные были предоставлены Конни Аэртс из Левенского университета. Предполагаемый значение для периода p составляет 0,161 день.

На рисунке также видно, что крайние дыры в данные можно обрабатывать с помощью P-сплайнов. С разницей- если они не были сохранены, параметрическая модель то, что приближается к сильному сглаживанию, будет отличается от полученного с помощью полиномиального агрессия. При оценке плотности инфляция дисперсии равна нежелательно. Работа нескольких первоклассных статисти- Об этом свидетельствуют граждане, в том числе и сам Джонс.

МНОГООБРАЗНЫЕ ПРИЛОЖЕНИЯ

Что касается многовариантных приложений, у нас есть эффективные Функции MATLAB для двумерных P-сплайнов, на основе тензорных произведений одномерных B- шлицы. В двух измерениях вероятность разреженного данные высоки. Чтобы предотвратить проблемы с идентификацией, штраф почти обязателен. Мы не оптимистичны. мистика по поводу обобщений в более высокие измерения. Теоретически это не так уж и сложно, но организация вычислений сложно. Также количество

Стр.32

Таблица 1

Потребительские испытания методов сглаживания; аббревиатуры означают: KS - ядровый сглаживатель; KSB, сглаживание ядра с биннингом; LR - локальная регрессия; LRB - локальная регрессия с биннингом; SS, сглаживающие шлицы; SSB, сглаживание сплайнов ленточным решателем; RSF, регресс- зубчатые шлицы с закрепленными узлами; RSA, регрессионные сплайны с адаптивными узлами; ПС, П-шлицы. Строка «Доступна адаптивная гибкость» означает что программная реализация легко доступна

Аспект	KS	KSB	LR	LRB	SS	SSB	RSF	ЮАР	PS
Скорость примерки	-	+	-	+	-	+	+	+	+
Скорость оптимизации	-	+	-	+	-	+	-	-	+
Граничные эффекты	-	-	+	+	+	+	+	+	+
Редкие дизайны	-	-	-	-	+	+	-	+	+
Полупараметрические модели	-	-	-	-	+	-	+	+	+
Ненормальные данные	+	+	+	+	+	+	+	+	+
Простая реализация	+	-	+	-	+	-	+	-	+
Параметрический предел	-	-	+	+	+	+	+	+	+
Специализированные лимиты	-	-	-	-	+	+	-	-	+
Инфляция дисперсии	-	-	+	+	+	+	+	+	+
Возможна адаптивная гибкость	+	+	+	+	+	+	-	+	+
Доступна адаптивная гибкость	-	-	-	-	-	-	-	+	-
Компактный результат	-	-	-	-	-	-	+	+	+
Сохранение моментов	-	-	+	+	+	+	+	+	+
Простые стандартные ошибки	-	-	+	+	-	+	+	+	+

базисные функции могут легко стать больше, чем количество наблюдений.

ШИРИНА БИНА ГИСТОГРАММ

Сэйн и Скотт хотели бы увидеть расследование того, как ширина бина гистограммы влияет на P- оценка плотности сплайна. Мы делаем это эмпирически с Старый Верный (это имя получает новое значение здесь) данные. На рисунке 9 мы построили расчетную плотность для пяти значений ширины бина: 0,2, 0,1, 0,05, 0,02 и 0,01. Оптимальное значение λ_{opt} было найдено путем попытки убывающая серия целых степеней 10, начало- на 10^4, останавливаясь, когда AIC начал расти. К последние три пары $\log \lambda$ AIC, парабола была подобрана; расположение его минимума дало $\log \lambda_{opt}$. Кажется что ячейки 0,2 слишком грубые, но для других значений При этом кривые практически не отличаются. Как правило большой палец мы могли бы получить гистограмму с 100 или несколько ящиков - хороший выбор.

ПОТРЕБИТЕЛЬСКИЙ ТЕСТ

Некоторые участники дискуссии сомневаются в том, что P-шлицы приблизиться к идеальной гладкости, как мы утверждаем. Каждое хорошее свойство, которое мы упомянули, также может быть найдено одним или несколькими другими методами. В таблице 1 мы разработали «потребительский тест» сглаживающих устройств на упростите сравнение. Мы пренебрегли большинством специальные изменения, которые были опубликованы для исправления такие проблемы, как граничные эффекты и разреженный дизайн, потому что они еще не появились в свободном доступе способное программное обеспечение. Значение «специализированных лимитов» будет объяснено ниже. Конечно можно поспорить о некоторых плюсах или минусах конкурентов- методов, но преимущества P-шлицев имеют прочная основа. Любой желающий может быть отслеживать за несколько часов на любом языке, поддерживающем матричные операции и / или регрессия, начиная с алгоритмы в нашем Приложении.

Рис. 9. Пять оценочных значений плотности для данных Old Faithful, основанные на на гистограммах с шириной бина 02, 01, 005, 002, 001; Кривая с самыми низкими пиками для ширины бункера 02

СЧАСТЛИВЫЙ КОНЕЦ

Да, мы считаем, что P-шлицы заслуживают места в в центре внимания. Они просты в использовании, легко обрабатываются. грамм и легко понять. Они уважают границы - Овен, нет проблем с разреженными рисунками и дают компактные результаты. Полиномиальный и экспоненциальный

Стр. 33

(синусоидальные) пределы могут быть установлены с помощью почти тривиальных изменяет оператор разницы в штрафах.

Тем не менее, еще многое предстоит сделать, особенно об оптимизации веса штрафа и о адаптивная гибкость. Лучшее понимание Требуется байесовская интерпретация штрафа. Мы продолжим наши исследования в этих областях. Мы надеюсь встретить там многих других, у которых также есть признал очарование P-шлицев.

БЛАГОДАРНОСТИ

Еще раз благодарим участников дискуссии за их в- вдохновляющие комментарии. Мы также признательны за совет (анонимных) судей и поддержка Ассо- младший редактор Дэвид Скотт (с особой благодарностью за или- обобщая обсуждение) и редактор Пол Свитцер о наш долгий путь к бумаге в печати.

ДОПОЛНИТЕЛЬНАЯ ЛИТЕРАТУРА

Кук, Р. Д. и Вайсберг, С. (1994). Графика регрессии. Wi- лей, Нью-Йорк.

Эйлерс, PHC (1988). Модели авторегрессии со скрытой вариабельностью способностей. В слушаниях COMPSTAT 1988 (Д. Эдвардс и Н.Э. Раун, ред.). Physica-Verlag.

Энгель, Дж. И Гассер, Т. (1995). Минимаксный результат для класса непараметрические оценки плотности. Непараметрическая статистика 4 327–334.

Обследование семейных расходов (1968–1983 гг.). Годовые базовые ленты и отчеты (1968–1983). Департамент занятости, Статистический отдел видение, Канцелярия Ее Величества, Лондон.

Фан, Дж. (1993). Сглаживатели локальной линейной регрессии и их минимальные imax эффективность. Аня. Статист. 21 196–216.

Fan, J. и Gijbels, I. (1995). Выбор полосы пропускания на основе данных в локальной полиномиальной аппроксимации: переменная полоса пропускания и пространственная приспособление. Дж. Рой. Статист. Soc. Сер. B 57 371–394.

Фан, Дж., Холл, П., Мартин, М.А., Патил, П. (1996). На ло- сглаживание непараметрических оценок кривых. J. Amer. Статист. Доц. 91 258–266.

Фоли, Дж. Д., Ван Дам, А., Файнер, С. К. и Хьюз, Дж. Ф. (1996). Компьютер Графика: Принципы а также Упражняться. Эддисон-Уэсли, Рединг, Массачусетс.

Фридман, JH (1991). Многомерные сплайны адаптивной регрессии (с обсуждением). Аня. Статист. 19 1–141.

Джонс, М.С., Маррон, Дж. С. и Шезер, С. Дж. (1996). A краткий обзор выбора полосы пропускания для оценки плотности. J. Amer. Статист. Доц. Появляться.

Кнейп, А. (1994). Заказал линейные сглаживания. Аня. Статист 22 835–866.

Куперберг К., Бозе С. и Стоун С.Дж. (1997). Полихото- мышечная регрессия. J. Amer. Статист. Доц. Появляться.

Куперберг К., Стоун С.Дж. и Чыонг Ю.К. (1995a). Haz- сильная регрессия. J. Amer. Статист. Доц. 90 78–94.

Куперберг К., Стоун С.Дж. и Чыонг Ю.К. (1995b). Логсплайновая оценка возможно смешанного спектрального распределения. тион. J. Time Ser. Анальный. 16 359–388.

Спекман, П.Л. (1983). Сглаживание сплайнов и оптимальные скорости сходимости в непараметрических регрессионных моделях. Аня. Статист. 13 970–983.

Стоун, С.Дж., Хансен, М., Куперберг, К. и Чыонг, Ю.К. (1996). Полиномиальные сплайны и их тензорные произведения в ex- склонен к линейному моделированию. Аня. Статист. Появляться.

Вахба, Г. (1978). Неправильные приоры, сглаживание шлицев и проблема защиты от ошибок модели в регрессии. Дж. Рой. Статист. Soc. Сер. B 40 364–372.