Ж. Антиль, А. Вайнберг-Аллен

D-оптимальное планирование для полиномиальной регрессии: выбор степени и робастность

Выбор степени регрессии является одной из основных проблем полиноминальной регрессии. В литературе по оптимальному планированию, как правило, предполагается, что статистическая модель известна. Однако на практике точная степень полиноминала достоверно не известна.

Более того, используемый план экспериментов может не соответствовать оптимальному, полученному на основе аналитического решения задачи. В настоящей работе мы обращаем внимание на то, что для D-оптимального планирования, отклонения от плана экспериментов гораздо менее важны, чем от модели. Таким образом, на основании D-оптимальности, мы предлагаем правило для выбора степени регрессии.

Мы также исследуем различные типы отклонений от модели для определения нового класса D-оптимальных планов экспериментов, который для анализируемых моделей устойчив и более эффективен, чем равномерные планы экспериментов.

Краткое изложение

Рассмотрим $E(y|x) = f^T(x)\beta$ — модель общей полиномиальной регрессии с гетероскедастичной ошибкой $V(y|x) = \sigma^2/\lambda(x)$, где $\lambda(x) > 0$ — функция эффективности эксперимента.

План эксперимента является вероятностной мерой ξ, определенной в области значений x. Оптимальный план минимизирует (или максимизирует) Φ_m, являющуюся некой выпуклой (или вогнутой) функцией информационной матрицы $M(\xi) = \int_I \lambda(x) f^T(x) f(x) d\xi(x)$.

$\Phi_m(M(\xi)) = [\mathrm{trace}(M(\xi))^{-1}]^{1/m}$ определяет набор классических критериев. При $m = 1$ мы получаем A-оптимальность. При $m \to \infty$ E-оптимальность определяется через собственные значения матрицы. D-оптимальность соответствует $m \to 0$.

D-оптимальные планы хорошо известны благодаря важному классу функций эффективности [Fedorov (1972)]. Антиль [Antille (1977)] предложил обобщенную версию результатов Федорова. Антиль, Дет и Вайнберг [Antille, Dette, Weinberg (2001)] предложили найденные аналитически D-оптимальные планы для функций эффективности $\lambda(x) = (1+x^2)^{\alpha+1} \exp(2\beta \arctan(x))$ и $\lambda(x) = x^{-\gamma} \exp(-\delta/x)$.

Более подробно об оптимальных планах эксперимента рассказывается в монографиях [Fedorov (1972)], [Silvey (1980)], [Atkinson, Donev (1992)] и [Pukelsheim (1993)].

Для сравнения планов в терминах *D*-оптимальности, мы определяем *D*-эффективность произвольного плана ξ относительно оптимального плана ξ^* как $D_{eff} = \frac{|M(\xi)|}{|M(\xi^*)|}$. Стандартизированная *D*-эффективность есть *n*-й корень *D*-эффективности, где *n* — степень полиномиальной регрессии.

Численные результаты, приведенные в различных графиках статьи, указывают на то, что возмущения граничных точек оказывают более значительное влияние на *D*-эффективность, чем возмущения любых других точек.

На практике точная степень полинома достоверно не известна. В этом разделе мы анализируем поведение детерминанта информационной матрицы относительно степени полиномиальной регрессии для классических функций эффективности. Для функций эффективности, удовлетворяющих $|\lambda(x)| \le 1$ для $x \in [-1;+1]$, стандартизированный детерминант информационной матрицы является убывающей функцией степени регрессии. Обратное верно для функций эффективности Лагерра. Таким образом, для *D*-оптимальных планов, генерируемых полиномами Лежандра и Якоби, обобщенная дисперсия является возрастающей функцией степени регрессии. Для функций эффективности Лагерра, или обобщенного Лагерра, верно обратное. В случае функции эффективности Эрмита общей закономерности не существует.

Таким образом, при анализе *D*-оптимальности можно сформулировать правило выбора степени регрессии. Для гомоскедастичных случаев и для функций эффективности типа Якоби степень регрессии должна быть «наименьшей приемлемой». Для функций эффективности Лагерра степень регрессии должна быть «наибольшей приемлемой». Для функции Эрмита каждый случай должен изучаться отдельно.

В качестве иллюстрации в табл. 1 приводятся стандартизированные детерминанты *D*-оптимальных планов для различных функций эффектиности.

Анализ, приведенный в этой части исследования, был мотивирован исследованиями Хьюберта [Huber (1975), (1981)], указавшего, что D-оптимальные планы крайне неробастны даже для достаточно небольших отклонений от линейности и показывавшего, что равномерные планы демонстрируют гораздо более удовлетворительное поведение, чем оптимальные.

Как видно из табл. 2, отклонение от модели в большей степени влияет на величину стандартизированного детерминанта, чем отклонения от оптимального плана. Мы видим также, что в равномерных планах с увеличением количества точек значения стандартизированных детерминантов также монотонно возрастают.

И, наконец, мы определяем новый класс планов (*D-оптимальные планы более высокого порядка*), который одновременно (1) более эффективен, чем равномерные планы, и (2) так же робастен как и равномерные планы и, следовательно, семейство *D*-оптимальных планов более высокого порядка может считаться почти оптимальным устойчивым семейством аппроксимированных планов, где, по определению, *D*-оптимальные планы более высокого порядка являются *D*-оптимальными планами порядка *k* для степени регрессии $l < k$.