Билет 20

Неточный метод Ньютона

Рассматривается задача безусловной оптимизации

Min x∈R^n

f(x),

где функция f является дважды непрерывно дифференцируемой.

Метод Ньютона строит последовательность точек x\_k, сходящуюся к решению задачи

Каждая итерация метода имеет вид:

x\_(k+1) = x\_k + αkdk.

Направление оптимизации dk находится из следующей системы линейных уравнений:

H\_k d\_k = −g\_k

где gk и Hk суть градиент и гессиан функции f в точке xk. Длина шага αk выбирается с помощью поиска по прямой.

В неточном методе Ньютона система решается с помощью метода сопряженных градиентов. В этом случае матрицу H\_k можно в явном виде в памяти не формировать, т. к. сама Hk методу не нужна; нужна только процедура умножения Hk на произвольный вектор d\_k. Кроме того, не имеет большого смысла решать систему сильно точно, если текущая точка xk находится далеко от оптимума. Поэтому обычно метод сопряженных градиентов останавливают, как только невязка

R\_k = H\_kd\_k + g\_k удовлетворяет условию

Kr\_k <= η\_k kg\_kk .

Неточный метод оптимизации Ньютона может применяться для поиска минимума функции ошибок. Функция ошибок существует как в задаче классификации так и в задаче регрессии.

В случае уравнения с одной переменной приближенные значения могут использоваться для решения уравнения для данной непрерывно дифференцируемой функции.

Квазиньютоновские методы — методы оптимизации, основанные на накоплении информации о кривизне целевой функции по наблюдениям за изменением градиента, чем принципиально отличаются от ньютоновских методов. Класс квазиньютоновских методов исключает явное формирование матрицы Гессе, заменяя её некоторым приближением. Это может существенно ускорить поиск экстремума и сократить затраты памяти, если имеется большое количество признаков.

Разложим градиент g^^-(x\_k) исходной в ряд Тейлора в окрестности точки очередного приближения x\_k^- по степеням следующего шага алгоритма s\_k^-:

g^-(x\_k^- + s\_k^-) приблизительно g^-(x\_k^-) + G(x\_k^-)s\_k^-

Тогда оценка матрицы Гессе B\_(k + 1) должна удовлетворять равенству:

B\_(k + 1) \* s\_k^- = y\_k^-

Где y\_k^- = g^-(x\_k^- + s\_k^-) – g(x\_k^-) – квазиньютоновское условие

На каждой итерации с помощью B\_k определяется следующее направление поиска p\_k^-, и матрица B обновляется с учётом вновь полученной информации о кривизне:

B\_k \* p\_k^- = -g^-(x\_k^-)

B\_(k + 1) = B\_k + U\_k

Где U\_k - матрица, характеризующая поправку, вносимую на очередном шаге.

В качестве начального приближения кладут единичную матрицу, таким образом первое направление p\_0^- будет в точности совпадать с направлением наискорейшего спуска.

Алгоритм DFP (Дэвидсон, Флетчер и Пауэлл, Davidson, Fletcher и Powell) — градиентный метод с положительно-определенной взвешивающей матрицей A\_s. Также, требует расчета только первых производных. Взвешивающая матрица A\_s рассчитывается рекурсивным способом.

A\_s = A\_(s – 1) + (delta\_(s – 1) \* delta\_(s – 1)’)/(delta\_(s – 1)’ \* gamma\_(s – 1)) + (A\_(s – 1) \* gamma\_(s – 1) \* gamma\_(s – 1)’ \* A\_(s – 1))/(gamma\_(s – 1)’ A\_(s – 1) gamma\_(s – 1))

можно увидеть, что матрица As будет положительно-определена в случае, если начальная матрица Л0 будет положительно-определена.

Процедура сходится быстро во многих статистических приложениях. В результате матрица А3 сходится к матрице Гессе —Н~1 — более предпочтительной с точки зрения теории. В принципе данный метод может позволять получить приближенную оценку обратной матрицы к матрице Гессе для расчета стандартных ошибок, без расчета второй производной или обратной матрицы. Однако, полученная оценка может быть довольно плохой на практике.

Уточненным алгоритмом DFP является алгоритм BFGS (Бойден, Флетчер, Голдфарб и Шеннон, Boyden, Fletcher, Goldfarb и Shannon), с

A\_s = A\_(s – 1) + (delta\_(s – 1) \* delta\_(s – 1)’)/(delta\_(s – 1)’ \* gamma\_(s – 1)) + (A\_(s – 1) \* gamma\_(s – 1) \* gamma\_(s – 1)’ \* A\_(s – 1))/(gamma\_(s – 1)’ A\_(s – 1) gamma\_(s – 1)) – (gamma\_(s – 1)’ A\_(s – 1) gamma\_(s – 1)) \* nu\_s – 1\*nu\_(s – 1)’)

LBFGS – улучшенный метод BFGS с ограничением на память. Не вычисляется вся матрица Гессе, что также позволяет снизить затраты на память и увеличить скорость.