Методы оптимизации в машинном обучении, ШАД, весна 2017

Семинар 3: Градиентный спуск

27 февраля 2017 г.

Внимание: Предварительная версия.

Задача 1. Пусть $\beta \geq 0$ и $f : \mathbb{R}^n \to \mathbb{R}$ — функция

$$f(x) := \frac{1}{2+\beta}\|x\|_2^{2+\beta}.$$

Рассмотрим градиентный спуск с постоянной длиной шага $\alpha > 0$ для минимизации функции f, запущенный из точки $x_0 \in \mathbb{R}^n \setminus \{0\}$. Определите, для каких длин шага α метод будет сходиться к минимуму $x^* = 0$. Какова при этом будет скорость сходимости (линейная/сублинейная)?

Решение. Найдем градиент:

$$\nabla f(x) = \|x\|_2^\beta x.$$

Значит, итерация метода записывается следующим образом:

$$x_{k+1} = x_k - \alpha\|x_k\|_2^\beta x_k = (1 - \alpha\|x_k\|_2^\beta)x_k.$$

Обозначим $z_k := \|x_k\|_2$. Тогда

$$z_{k+1} = |1 - \alpha z_k^\beta| z_k. \tag{1}$$

Рассмотрим три возможные ситуации:

1. Длина шага α очень большая: $\alpha > 2z_0^{-\beta}$. В этом случае нетрудно понять, что z_{k+1} будет расходиться как минимум со скоростью геометрической прогрессии.

2. Длина шага $\alpha = 2z_0^{-\beta}$. В этом случае метод будет стоять на месте: $x_0 = x_1 = \dots$.

3. Длина шага $\alpha < 2z_0^{-\beta}$. Тогда будет монотонное убывание расстояний: $z_{k+1} \leq z_k$. Поскольку последовательность z_k ограничена снизу, то она имеет предел. Покажем, что этот предел в точности равен нулю. Обозначим предел через a. Пусть $a > 0$. Тогда, переходя в (1) к пределу, получим

   $$a = |1 - \alpha a^\beta| a.$$

   Поскольку $a \neq 0$, то отсюда

   $$|1 - \alpha a^\beta| = 1.$$

   Поскольку $\alpha > 0$ и $a > 0$, то

   $$\alpha a^\beta = 2 \qquad \Leftrightarrow \qquad \alpha = 2a^{-\beta}.$$

   Заметим, что это невозможно, поскольку $2z_0^{-\beta} < \alpha$ и z_k монотонно убывает. Итак, $a = 0$.

Заметим, что сходимость есть только в последнем случае, когда $\alpha < 2z_0^{-\beta}$. Скорость сходимости будет сублинейной, поскольку

$$\frac{z_{k+1}}{z_k} = |1 - \alpha z_k^\beta| \to 1.$$

(Аналогичная скорость сходимости будет и по функции.)

Задача 2. Пусть $f : \mathbb{R}^n \to \mathbb{R}$ — функция

$$f(x) = \frac{2}{3}\|x\|^{3/2}.$$

Заметим, что эта функция является непрерывно дифференцируемой, однако не имеет липшицев градиент (в окрестности нуля). Рассмотрите поведение градиентного спуска на этой функции с постоянным шагом $\alpha > 0$, запущенного из точки $x_0 \in \mathbb{R}^n \setminus \{0\}$.

Решение. Посчитаем градиент:

$$\nabla f(x) = \|x\|_2^{-1/2} x.$$

Тогда итерация градиентного спуска:

$$x_{k+1} = x_k - \alpha\|x_k\|_2^{-1/2} x_k = (1 - \alpha\|x_k\|_2^{-1/2}) x_k.$$

Перейдем к $z_k := \|x_k\|_2$:

$$z_{k+1} = |1 - \alpha z_k^{-1/2}| z_k.$$

Отсюда видно, что если $\alpha < 2z_k^{1/2}$, то $z_1 < z_0$. Однако, это означает, что $z_2 > z_2$. А, значит, $z_3 < z_2$. Получается коллебательный процесс. Таким образом, «если повезет», то метод сойдется, а иначе будет бесконечно коллебаться. (Действительно, метод остановится, только если z_k в какой-то момент в точности станет равно нулю.)

Задача 3 (Градиентный спуск и седловые точки)**.** Пусть $A \in \mathbb{S}^n$ — симметричная невырожденная матрица, имеющая хотя бы одно строго положительное и хотя бы одно строго отрицательное значение ($n \geq 2$). Рассмотрите градиентный спуск с постоянной длиной шага $\alpha > 0$, запущенный из точки $x_0 \in \mathbb{R}^n \setminus \{0\}$. Определите, как будет вести себя метод в зависимости от α и x_0.

Решение. В этом случае единственная стационарная точка $x^* = 0$ является седловой.

Градиентный спуск:

$$x_{k+1} = x_k - \alpha A x_k = (I_n - \alpha A) x_k.$$

Значит,

$$x_k = (I_n - \alpha A)^k x_0.$$

Рассмотрим спектральное разложение $A = Q\Lambda Q^T$, где $\Lambda := \mathrm{Diag}\{\lambda_1, \ldots, \lambda_n\}$ и $\lambda_1 \geq \cdots \geq \lambda_s > 0 > \lambda_{s+1} \geq \ldots \lambda_n$. Обозначим $\tilde{x}_k := Q^T x_k$. Тогда

$$\tilde{x}_k = (I_n - \alpha\Lambda)^k \tilde{x}_0.$$

В координатах:

$$\tilde{x}_{k,i} = (1 - \alpha\lambda_i)^k \tilde{x}_{0,i}.$$

Заметим, что для $\lambda_i < 0$ будет расходимость, кроме тех случаев, когда $\tilde{x}_{0,i} = 0$. То есть если x_0 *не* лежит в подпространстве, отвечающему отрицательным собственным значениям, то x_k расходится. Если же x_0 лежит в соответсвующем подпространстве, то сходимость к седловой точке $x^* = 0$.