Гапеев Д.А.
РК6-51Б

Задача 2.11 по вычислительной математике

Пусть $\theta \in \mathbb{R}^d$ - точка в многомерном пространстве, и пусть E - целевая функция. Для оптимизации целевой функции необходима производная $\frac{\partial}{\partial\theta}E(\theta)$, аналитическое выражение для которой неизвестно, но можно поступить следующим образом, создав n векторов $\epsilon_1, ..., \epsilon_n \in \mathbb{R}^d$ случайных величин, распределённых по стандартному нормальному закону ($\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2),\ i \in [1..n],\ j \in [1..d]$), и измерить значение производной E в точках $\theta + \epsilon_1, ..., \theta + \epsilon_n$. По полученным значениям функции E можно найти приближённое значение производной E в точке θ, используя формулу

$$E'(\theta) \approx \widehat{E}'(\theta) = a \cdot \mathrm{mean}(b \cdot E(\theta + \epsilon_1), ..., b \cdot E(\theta + \epsilon_n)), \tag{1}$$

где mean - среднее значение, которое приблизительно равно математическому ожиданию случайной величине, а a и b - некоторые векторы с неизвестными значениями элементов.

Необходимо:

1. Вывести значения a и b в приведённой выше формуле, пользуясь готовыми формулами моментов случайной величины, распределённой по нормальному закону :

$$\mathbb{M}(\epsilon) = 0,$$
$$\mathbb{M}(\epsilon^2) = \sigma^2,$$
$$\mathbb{M}(\epsilon^3) = 0,$$
$$\mathbb{M}(\epsilon^4) = 3\sigma^4,$$

2. Вывести выражение для оценки погрешностей производной в терминах метрики среднеквадратичной ошибки (MSE):

$$MSE = \frac{1}{d}\sum_{i=1}^{d}\left[\left(E_i'(\theta) - \widehat{E}_i'(\theta)\right)^2\right] \tag{2}$$

3. Ответить на вопрос, зависит ли погрешность от размерности пространства.

4. Объяснить, почему на практике происходит снижение точности при увеличении размерности пространства, когда число генерируемых векторов n фиксировано (даже если ваш ответ на предыдущий вопрос доказывает, что ошибка не зависит от размерности пространства).

5. Найти оценку верхней границы для погрешности приведённой выше формулы для функции Растригина на интервале $[-5.12, 5.12]$:

$$E(\theta) = 10d + \sum_{i=1}^{d} \Big[\theta_i^2 - 10\cos(2\pi\theta_i)\Big] \tag{3}$$

6. Опираясь на те же соображения, с помощью которых были получены вектора a и b, вывести формулу для определения приблизительного значения второй производной.

Решение:

1. Для нахождения значений a и b разложим целевую функцию E в ряд Тейлора:

$$E(\theta+\epsilon) = E(\theta) + E'(\theta)\epsilon + \frac{E''(\theta)\epsilon^2}{2} + O(\epsilon^3) \tag{4}$$

$$E(\theta+\epsilon)\epsilon = E(\theta)\epsilon + E'(\theta)\epsilon^2 + \frac{E''(\theta)\epsilon^3}{2} + O(\epsilon^4) \tag{5}$$

Отбросим $O(\epsilon^4)$ и возьмем мат.ожидание от обоих сторон равенства:

$$\mathbb{E}_{\epsilon\sim\mathcal{N}(0,\sigma^2)}\Big[E(\theta+\epsilon)\epsilon\Big] \approx \mathbb{E}_{\epsilon\sim\mathcal{N}(0,\sigma^2)}\Big[E(\theta)\epsilon + E'(\theta)\epsilon^2 + \frac{E''(\theta)\epsilon^3}{2}\Big] \tag{6}$$

Так как $\mathbb{M}(\epsilon) = \mathbb{M}(\epsilon^3) = 0$ и $\mathbb{M}(\epsilon^2) = \sigma^2$, тогда

$$\mathbb{E}_{\epsilon\sim\mathcal{N}(0,\sigma^2)}\Big[E(\theta+\epsilon)\epsilon\Big] = E'(\theta)\cdot\sigma^2 \tag{7}$$

$$E'(\theta) = \frac{1}{\sigma^2}\cdot\mathbb{E}_{\epsilon\sim\mathcal{N}(0,\sigma^2)}\Big[\epsilon E(\theta+\epsilon)\Big] \tag{8}$$

Сравнивая полученную формулу с формулой (1) можно сделать вывод, что:

$$a = \frac{1}{\sigma^2}\ ;\ b = \epsilon_i$$

2. Для оценки погрешностей производной в терминах метрики среднеквадратической ошибки(MSE), выведем определение MSE для каждой компоненты $E'(\theta)$:

$$MSE_i = \left[\Big(E_i'(\theta) - \widehat{E}_i'(\theta)\Big)^2\right], \tag{9}$$

где MSE_i - среднеквадратическая ошибка i-ой компоненты производной, $E_i'(\theta)$ - значение i-ой компоненты производной в точке θ, $\widehat{E}_i'(\theta)$ - оценка i-ой компоненты производной, полученная при использовании формулы (1). Выразим общую среднеквадратическую ошибку MSE для всех компонент производной:

$$MSE = \frac{1}{d}\sum_{i=1}^{d}\Big[MSE_i\Big], \tag{10}$$

Тогда выражение для оценки погрешности производной в терминах MSE:

$$MSE = \frac{1}{d}\sum_{i=1}^{d}\left[\Big(E_i'(\theta) - \widehat{E}_i'(\theta)\Big)^2\right]$$

3. Выполним вывод общего вида погрешности. Для этого в формуле (5) запишем $O(\epsilon^n)$ в виде остаточного члена : $\frac{E^{(n)}(\xi)\epsilon^m}{n!}$. Затем возьмём мат.ожидание от этого остаточного члена(также как в формуле 6):

$$\mathbb{E}(\frac{E^{(n)}(\xi)\epsilon^m}{n!}) = E^{(n)}(\xi)\cdot a\cdot \sigma^k, \tag{11}$$

где a - число, $E^{(n)}(\xi)$ - n-ая производная функции E, σ^k - гиперпараметр (выбирается перед запуском алгоритма). Тогда для оценки погрешности необходимо будет вычислять $E^{(n)}(\xi)$, для этого необходимо найти $E_i^{(n)}(\theta)$ по каждой компоненте i и затем сложить. Исходя из этого, можно сделать вывод, что погрешность зависит от размерности пространства.

4. Проанализировав выражение для погрешности в общем виде, можно сделать вывод, что при увеличении d, уменьшается точность, так как увеличивается количество компонент и соответственно увеличивается значение $E^{(n)}(\xi)$.

Также при увеличении размерности пространства увеличивается вычислительная сложность вычисления производных, так как для каждой точки требуется выполнить оценку функции E в окрестности точки θ. Это может быть вычислительно затратно и времязатратно. То есть, возрастает вычислительная погрешность алгоритма. Увеличение размерности также означает, что точки, сгенерированные для оценки производной, становятся более разреженными в пространстве, и это может привести к менее точной оценке производной.

Таким образом, на практике при увеличении размерности пространства, может потребоваться выбирать более большое n, чтобы компенсировать увеличение размерности и улучшить точность оценки производной.

5. Для нахождения оценки верхней границы погрешности для функции Растригина на интервале $[-5.12, 5.12]$ воспользуемся формулой (11):

$$E^{(n)}(\xi)\cdot a\cdot \sigma^k$$

Для нахождения n, a, k расммотрим формулу (5), в которой заменим $O(\epsilon^4)$ на остаточный член:

$$\frac{E'''(\xi)\epsilon^4}{3!}$$

Затем по формуле (6) возьмем матожидание от данной величины и получим:

$$\mathbb{E}(\frac{E'''(\xi)\epsilon^4}{3!}) = E'''(\xi) \cdot \frac{1}{2} \cdot \sigma^4, \tag{12}$$

Так как необходимо найти верхнюю границу погрешности, то найдем максимально возможное значение $E'''(\xi)$, где $\xi \in [-5.12, 5.12]$. Для нахождения $E'''(\xi)$ в многомерном пространстве, нам сначала потребуется вычислить третьи производные по каждой компоненте ξ_i и затем оценить их на указанном интервале. Поскольку компоненты ξ_i независимы, мы можем рассмотреть одну компоненту и оценить её третью производную.

$$E'(\xi_i) = 2\xi_i + 20\pi \sin(2\pi\xi_i) \tag{13}$$

$$E''(\xi_i) = 2 + 40\pi^2 \cos(2\pi\xi_i) \tag{14}$$

$$E'''(\xi_i) = -80\pi^3 \sin(2\pi\xi_i) \tag{15}$$

Для максимума второй производной $E'''(\xi)$ на интервале, нам нужно рассмотреть максимальное значение по модулю для синуса $\sin(2\pi\xi_i)$, которое равно 1(при $\xi_i = \frac{1}{4}, \frac{5}{4}, ...$) или же -1(при $\xi_i = \frac{3}{4}, \frac{7}{4}, ...$) . Таким образом, максимальное значение по модулю третьей производной $E'''(\xi)$ будет равно:

$$E'''_{max}(\xi_i) = |-80\pi^3| = 80\pi^3 \tag{16}$$

Чтобы найти максимальное значение по модулю третьей производной по всем компонентам для функции $E(\xi)$, умножим полученное значение для одной компоненты на размерность d:

$$E'''_{max}(\xi) = 80\pi^3 \cdot d \tag{17}$$

Пусть $\sigma^4 = 1$, тогда оценка верхней границы для погрешности формулы (1) для функции Растригина на интервале $[-5.12, 5.12]$:

$$E'''_{max}(\xi) \cdot \frac{1}{2} \cdot \sigma^4 = 80\pi^3 \cdot d \cdot \frac{1}{2} \cdot 1 = 40\pi^3 \cdot d \tag{18}$$

6. Выведем формулу для определения приблизительного значения второй производной. Для этого вспомним разложение целевой функции E в ряд Тейлора:

$$E(\theta+\epsilon) = E(\theta) + E'(\theta)\epsilon + \frac{E''(\theta)\epsilon^2}{2} + O(\epsilon^3)$$

Отбросим $O(\epsilon^3)$ и возьмем мат.ожидание от обоих сторон равенства:

$$\mathbb{E}_{\epsilon\sim\mathcal{N}(0,\sigma^2)}\Big[E(\theta+\epsilon)\Big] \approx E(\theta) + \frac{E''(\theta)\sigma^2}{2} \tag{19}$$

Далее вновь используем разложение в ряд Тейлора, при это умножив обе стороны равенства на $\frac{\epsilon^2}{\sigma^2}$ и сразу отбросим $O(\epsilon^3)$:

$$\frac{\epsilon^2}{\sigma^2}E(\theta+\epsilon) = \frac{\epsilon^2}{\sigma^2}E(\theta) + E'(\theta)\frac{\epsilon^3}{\sigma^2} + \frac{E''(\theta)\epsilon^4}{2\sigma^2} \tag{20}$$

Возьмем мат.ожидание от обоих сторон равенства:

$$\mathbb{E}_{\epsilon\sim\mathcal{N}(0,\sigma^2)}\Big[\frac{\epsilon^2}{\sigma^2}E(\theta+\epsilon)\Big] \approx E(\theta) + \frac{3E''(\theta)\sigma^2}{2} \tag{21}$$

Вычтем формулу (19) из формулы (21):

$$\mathbb{E}_{\epsilon\sim\mathcal{N}(0,\sigma^2)}\Big[\Big(\frac{\epsilon^2}{\sigma^2}-1\Big)E(\theta+\epsilon)\Big] = E''(\theta)\cdot\sigma^2 \tag{22}$$

$$E''(\theta) = \frac{1}{\sigma^2}\cdot\mathbb{E}_{\epsilon\sim\mathcal{N}(0,\sigma^2)}\Big[\Big(\frac{\epsilon^2}{\sigma^2}-1\Big)E(\theta+\epsilon)\Big] \tag{23}$$

$$E''(\theta) \approx \widehat{E}''(\theta) = \frac{1}{\sigma^2}\cdot\mathrm{mean}\Big[\Big(\frac{\epsilon_1^2}{\sigma^2}-1\Big)E(\theta+\epsilon_1), ..., \Big(\frac{\epsilon_n^2}{\sigma^2}-1\Big)E(\theta+\epsilon_n)\Big] \tag{24}$$

Выражение (24) является формулой для определения приблизительного значения второй производной.