Метод скорейшего спуска (градиентного спуска)

Задача: минимизация функции $f(t_{1,} t_{2,} \dots t_k)$
Задача: найти $argmin(f(t_{1,} t_{2,} \dots t_k))$

Почему не работает

$$\frac{\partial f(t_1, t_2, \dots, t_k)}{\partial t_1} = 0$$
$$\dots$$
$$\frac{\partial f(t_1, t_2, \dots, t_k)}{\partial t_k} = 0$$

Временно рассмотрим случай функции одного аргумента

$f(t)$

t_0 - начальное приближение
λ - скорость обучения (learning rate)

$$t_{i+1} = t_i - \lambda \cdot f'(t_i)$$

Если λ слишком мал?
Если λ слишком велик?
Если функция $f(t)$ имеет плато?

В ходе минимизации λ должен убывать

Локальный минимум.

Вернемся к минимизации функции $f(t_{1,} t_{2,} \dots t_k)$

- начальное приближение точка $(t_{1,0}, t_{2,0}, \dots t_{k,0})$

$$t_{1,i+1} = t_{1,i} - \lambda \cdot \left. \frac{\partial f(t_1, t_2, \dots, t_k)}{\partial t_1} \right|_{t_{1,i}, t_{2,i}, \dots, t_{k,i}}$$
$$\dots$$
$$t_{k,i+1} = t_{k,i} - \lambda \cdot \left. \frac{\partial f(t_1, t_2, \dots, t_k)}{\partial t_k} \right|_{t_{1,i}, t_{2,i}, \dots, t_{k,i}}$$

Оценки параметров регрессионной модели методом скорейшего спуска.

Критерий качества $Q=\sum_{i=1}^{n} \epsilon_i^2$

Вектор частных производных (градиент) критерия качества

$$\nabla_\beta Q = \begin{pmatrix} \frac{\partial Q}{\partial \beta_0} \\ \frac{\partial Q}{\partial \beta_1} \\ \cdots \\ \frac{\partial Q}{\partial \beta_k} \end{pmatrix} = 2 \cdot X^T \cdot (X \cdot \beta - y)$$

$$\beta^{(i+1)} = \beta^{(i)} - \lambda \cdot \nabla_\beta Q \Big|_{\beta^{(i)}}$$

Проверьте себя: где векторы, где скаляры?

Правила остановки
в случае регрессии, например $max(\nabla_\beta Q) < \epsilon$

Batch
minibatch
Stochastic Gradient Descent

Нестеров
Adam