

Стохастический градиентный спуск

Евгений Соколов

Градиентный спуск

Инициализация: $w^0 = 0$

Цикл по $t = 1, 2, 3, \dots$:

$$w^t = w^{t-1} - \eta_t \nabla Q(w, X)$$

Если $\|w^t - w^{t-1}\| < \varepsilon$, то завершить

Градиент функционала

$$\nabla_w Q(w, X) = \frac{2}{\ell} X^T (Xw - y)$$

$$\frac{\partial Q}{\partial w_j} = \frac{2}{\ell} \sum_{i=1}^{\ell} x_i^j (\langle w, x_i \rangle - y_i)$$

Градиент функционала

$$\nabla_w Q(w, X) = \frac{2}{\ell} X^T (Xw - y)$$

$$\frac{\partial Q}{\partial w_j} = \frac{2}{\ell} \sum_{i=1}^{\ell} x_i^j (\langle w, x_i \rangle - y_i)$$

Суммирование по всей выборке!

Градиент функционала

$$\nabla_w Q(w, X) = \frac{2}{\ell} X^T (Xw - y)$$

$$\frac{\partial Q}{\partial w_j} = \frac{2}{\ell} \sum_{i=1}^{\ell} x_i^j (\langle w, x_i \rangle - y_i)$$

Как поменять веса, чтобы
улучшить качество на объекте x_i

Градиент функционала

$$\nabla_w Q(w, X) = \frac{2}{\ell} X^T (Xw - y)$$

$$\frac{\partial Q}{\partial w_j} = \frac{2}{\ell} \sum_{i=1}^{\ell} x_i^j (\langle w, x_i \rangle - y_i)$$

Как поменять веса, чтобы улучшить
качество на всей выборке

Стохастический градиентный спуск (SGD)

Инициализация: $w^0 = 0$

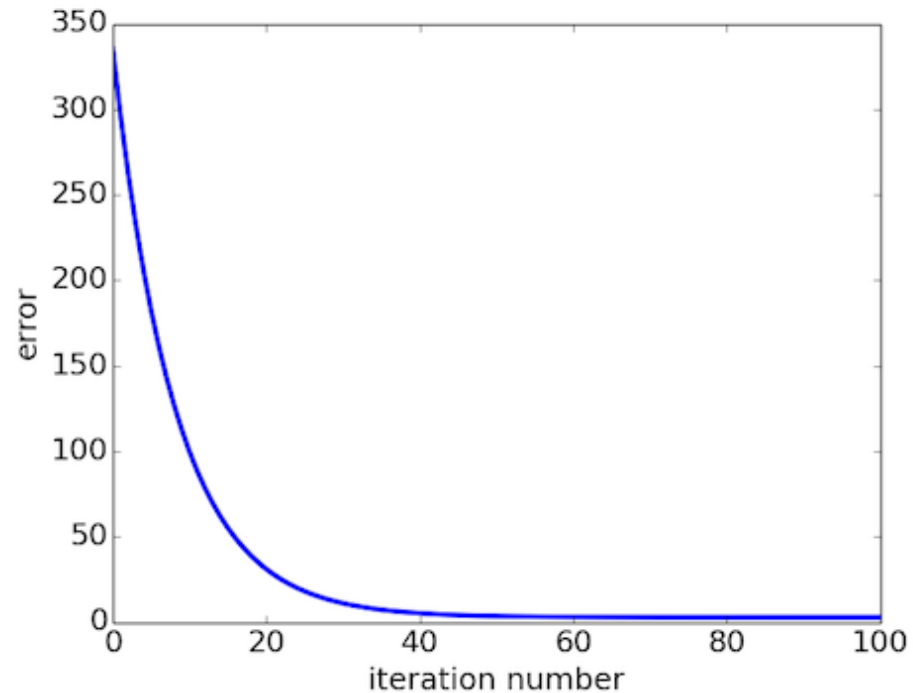
Цикл по $t = 1, 2, 3, \dots$:

 выбрать случайный объект x_i из X

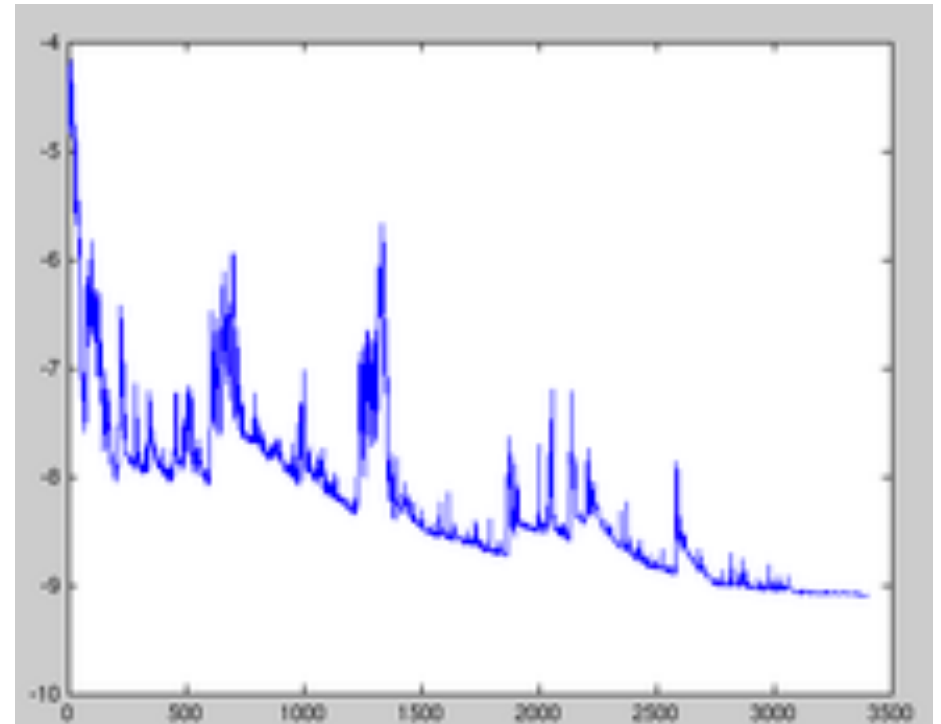
$$w^t = w^{t-1} - \eta_t \nabla Q(w, \{x_i\})$$

Если $\|w^t - w^{t-1}\| < \varepsilon$, то завершить

Сходимость



Градиентный спуск



Стохастический
градиентный спуск

Преимущества SGD

- Быстрее выполняется один шаг
- Не требует хранения выборки в памяти
- Подходит для онлайн-обучения

Резюме

- Градиентный спуск требует вычисления полного градиента
- Стохастический градиентный спуск использует лишь один объект
- SGD позволяет обучать алгоритм на больших выборках
- В следующем видео: линейный классификатор