

## Задача минимизации

$$\min_{x \in \mathbb{R}^d} \left[ f(x) \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^n f_i(x) \right]$$

Сильная выпуклость и  
Липшицевы градиенты

$$\phi(x) \geq \phi(y) + \langle \nabla \phi, x - y \rangle + \frac{\mu}{2} \|x - y\|^2$$

$$\|\nabla^2 \phi(x) - \nabla^2 \phi(y)\| \leq \|x - y\|$$

## Стохастический метод Ньютона

**Initialize:** Choose starting iterates  $w_1^0, w_2^0, \dots, w_n^0 \in \mathbb{R}^d$

**for**  $k = 0, 1, 2, \dots$  **do**

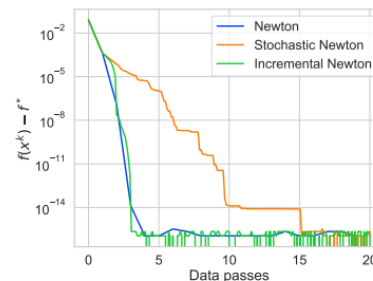
$$x^{k+1} = \left[ \frac{1}{n} \sum_{i=1}^n \nabla^2 f_i(w_i^k) \right]^{-1} \left[ \frac{1}{n} \sum_{i=1}^n \nabla^2 f_i(w_i^k) w_i^k - \nabla f_i(w_i^k) \right]$$

Choose a subset  $S^k \subseteq \{1, 2, \dots, n\}$

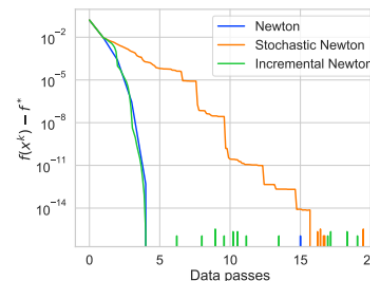
$$w_i^{k+1} = \begin{cases} x^{k+1} & i \in S^k \\ w_i^k & i \notin S^k \end{cases}$$

**end for**

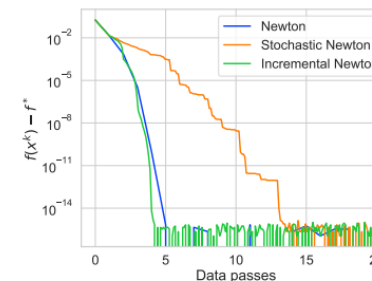
## Базовая версия алгоритма



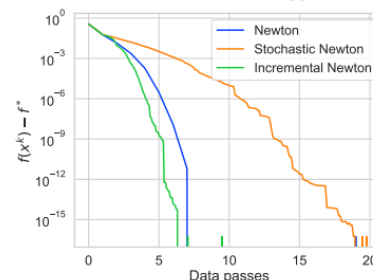
(a) Madelon,  $\lambda = \frac{1}{100n}$



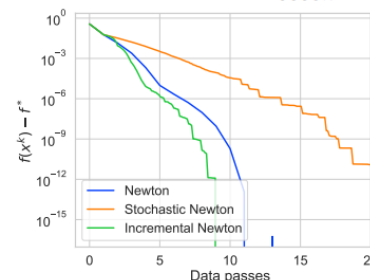
(b) Madelon,  $\lambda = \frac{1}{10000n}$



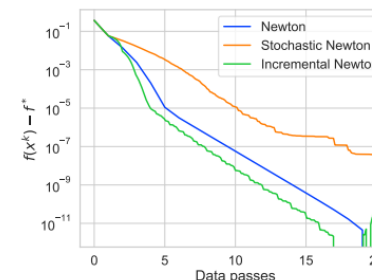
(c) Madelon,  $\lambda = 0$



(d) a8a,  $\lambda = \frac{1}{100n}$



(e) a8a,  $\lambda = \frac{1}{10000n}$



(f) a8a,  $\lambda = 0$

## Мотивация

Для SGD хорошо работают разные стратегии семплинга

Пример: Importance sampling – класть в  $S^k$  векторы  $w_i$  в зависимости от гладкости  $f_i$

## Цель проекта

Использовать разные стратегии для Стохастического метода Ньютона