

Incremental Newton Method for Minimizing Big Sums of Functions

НИС Машинное обучение и приложения, 2016/2017

1. В чем состоит основная идея метода Ньютона?
2. Объяснить интуитивно, что такое линейная/сублинейная/суперлинейная скорости сходимости.
3. Какую скорость сходимости имеют методы градиентного спуска и Ньютона на сильно выпуклых функциях? Для любой ли начальной точки x_0 эти методы гарантируют сходимость?
4. В чем состоят преимущества и недостатки метода Ньютона по сравнению с методом градиентного спуска (на сильно выпуклых функциях)?
5. Что мешает применять методы градиентного спуска (и/или Ньютона) в явном виде для задач минимизации вида

$$\frac{1}{n} \sum_i f_i(x) \rightarrow \min_{x \in \mathbb{R}^d} ?$$

6. В чем преимущества и недостатки стохастического градиентного метода по сравнению с градиентным спуском (для минимизации сильно выпуклой функции $(1/n) \sum_i f_i(x), x \in \mathbb{R}^d$)?
7. Пусть имеется сильно выпуклая задача оптимизации $(1/n) \sum_i f_i(x) \rightarrow \min_{x \in \mathbb{R}^d}$, и требуется решить эту задачу с точностью ϵ . В зависимости от ϵ , какой из двух методов — градиентный спуск или стохастический градиентный метод — следует применять и почему?