Московский государственный университет имени М.В.Ломоносова
Факультет вычислительной математики и кибернетики
Кафедра математических методов прогнозирования

Отчет по заданию

Минимизация суммы функций

Выполнил студент 517 группы

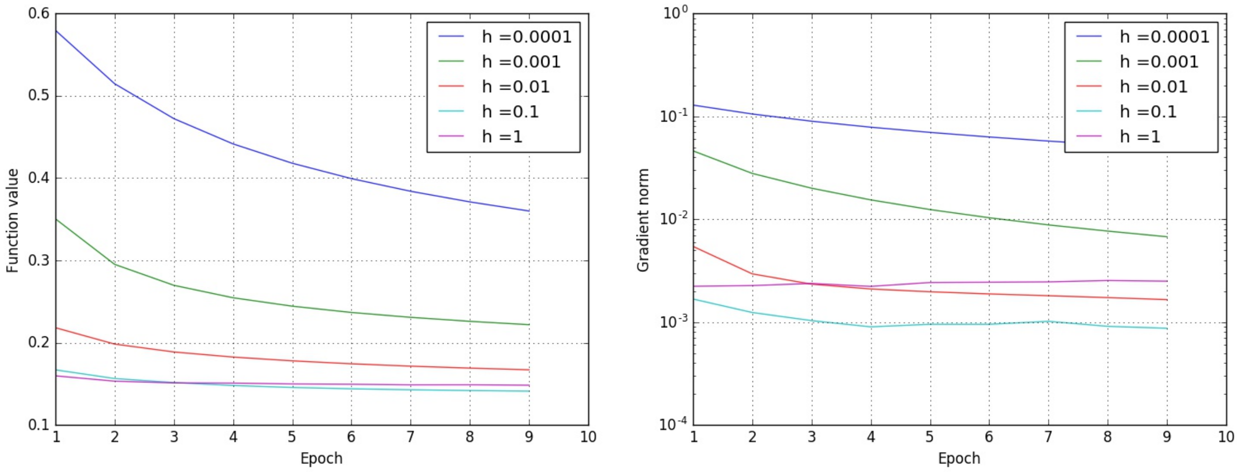
Камалов Руслан Рамилевич

Москва
2016

1 Тестирование на датасете w5a

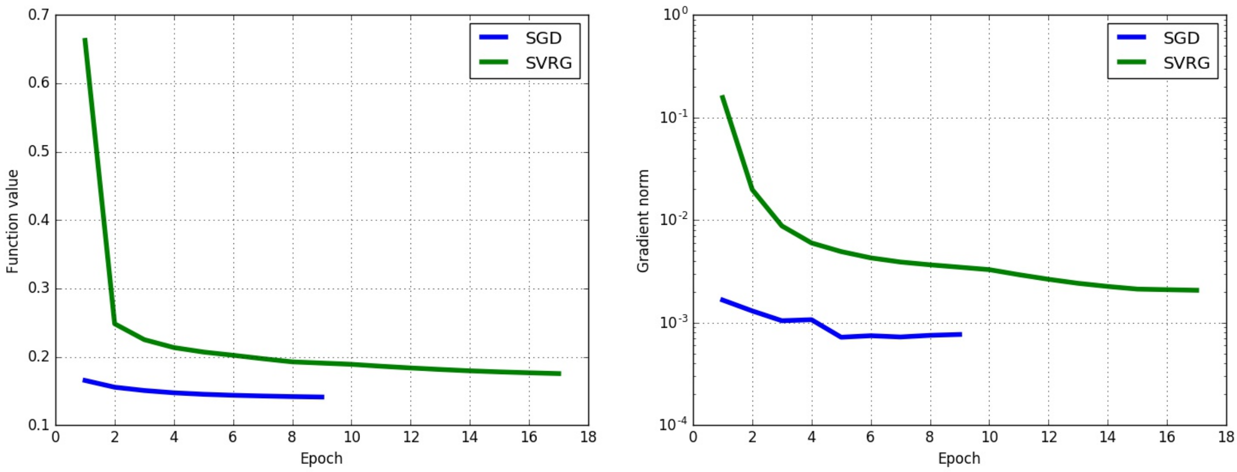
Модель: - логистическая регрессия
Датасет: - w5a

Работа метода SGD при различных длинах шага



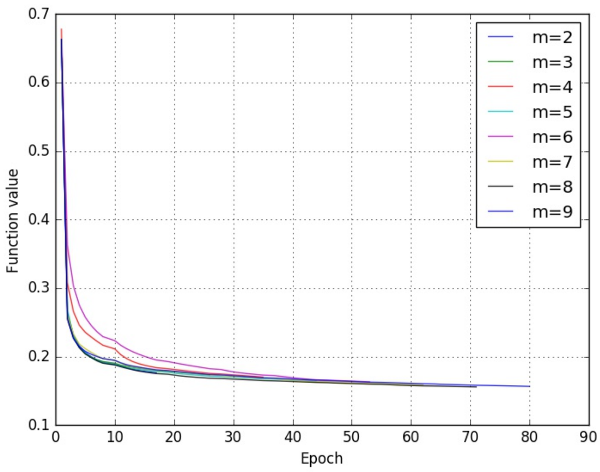
Решаемая задача является выпуклой и результаты эксперимента согласуются с теоретическими результатами относительно сходимости метода на выпуклых задачах. Также можно сделать вывод, что не стоить выбирать шаг слишком большим или наоборот слишком маленьким. Лучший результат показал SGD с шагом $h = 0.1$ как по значению функции, так и по значению градиента. При выборе слишком большого шага $h = 1$, метод практически стоит на месте, а при выборе слишком маленького $h = 0.00001$ сходиться очень медленно.

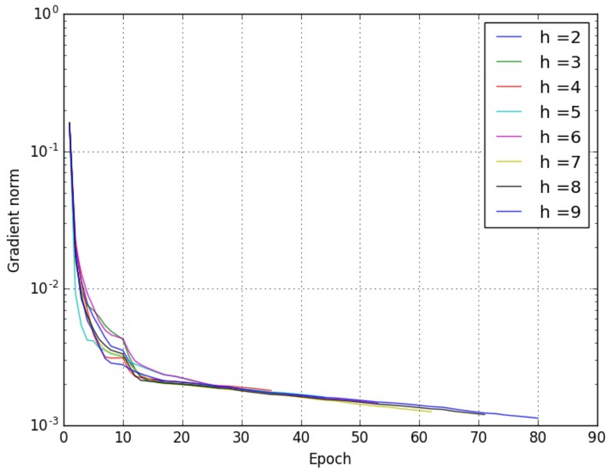
Сравнение методов SGD с лучшей длиной шага и SVRG



Метод SGD показал лучший результат на обоих графиках, но стоит отметить, что оба метода резко снижают скорость сходимости после 1-2 первых эпох.

Работа метода SVRG в зависимости от длины внутреннего цикла



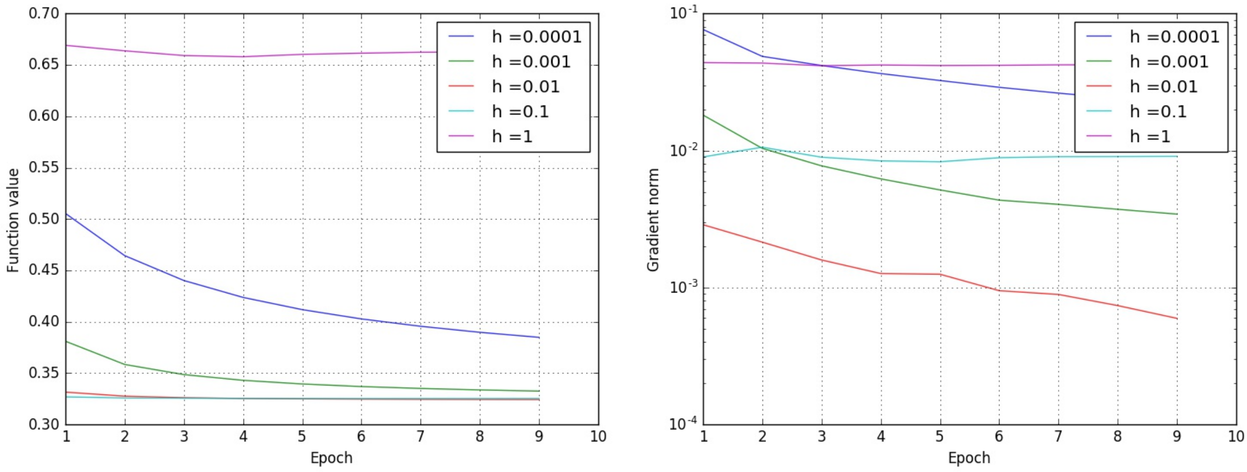


На обоих графиках видно, что метод работает тем лучше, чем длиннее внутренний цикл. Минимальную точность показал метод с $m = 2$, максимальную $m = 9$. Следовательно чем точнее хотим решить задачу, тем длиннее нужно выбирать длину внутреннего цикла. Процедура адаптивного подбора константы Липшица в методе SVRG на этом примере показала себя хорошо – график сходимости монотонно убывает.

2 Тестирование на датасете a9a

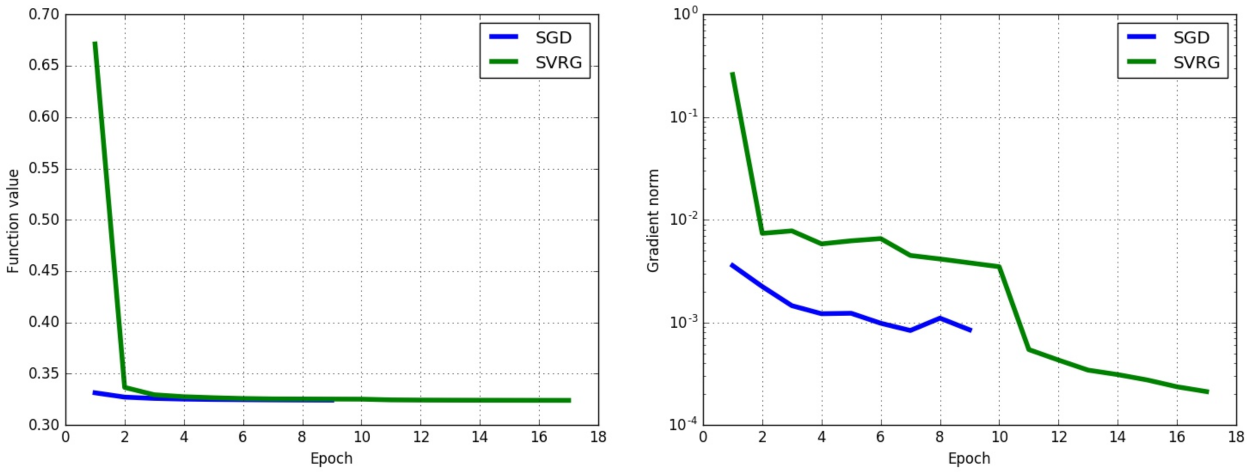
Модель: - логистическая регрессия
Датасет: - a9a

Работа метода SGD при различных длинах шага



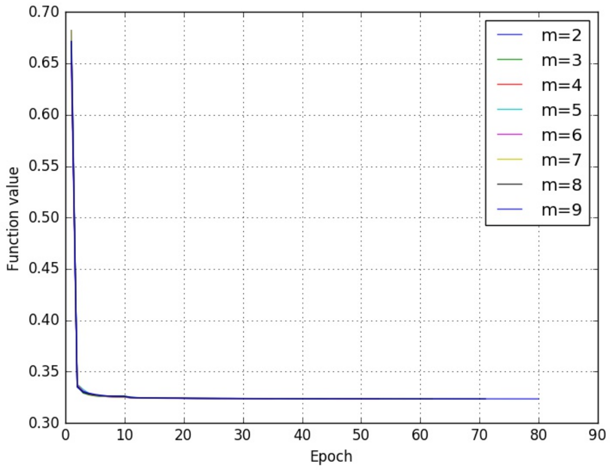
Графики сходимости очень похожи на аналогичные из предыдущего эксперимента. Обе эти задачи выпуклые, поэтому метод должен сходиться. Оптимальная длина шага в этом эксперименте $h = 0.01$.

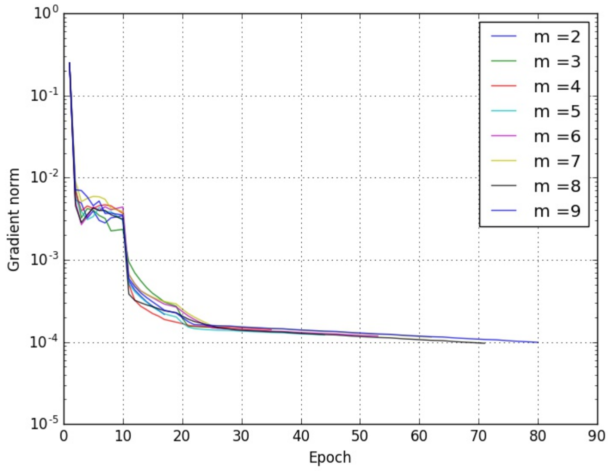
Сравнение методов SGD с лучшей длиной шага и SVRG



В этой задаче лучше себя показал метод SVRG,

Работа метода SVRG в зависимости от длины внутреннего цикла



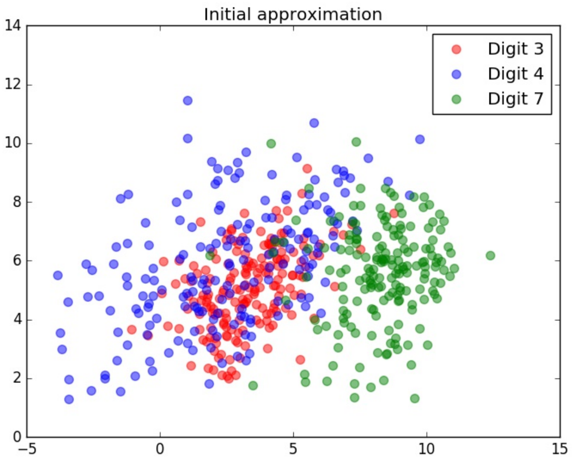


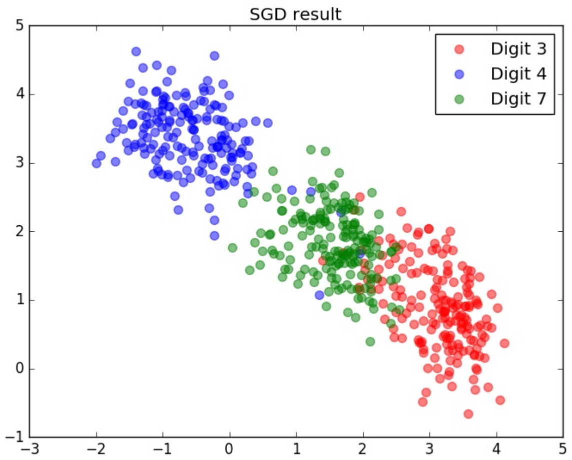
Графики схожи с аналогичными прошлого предыдущего эксперимента. Адаптивный подбор по схеме Нестерова сработал хорошо на этом примере.

3 Тестирование на датасете digits, архитектура I

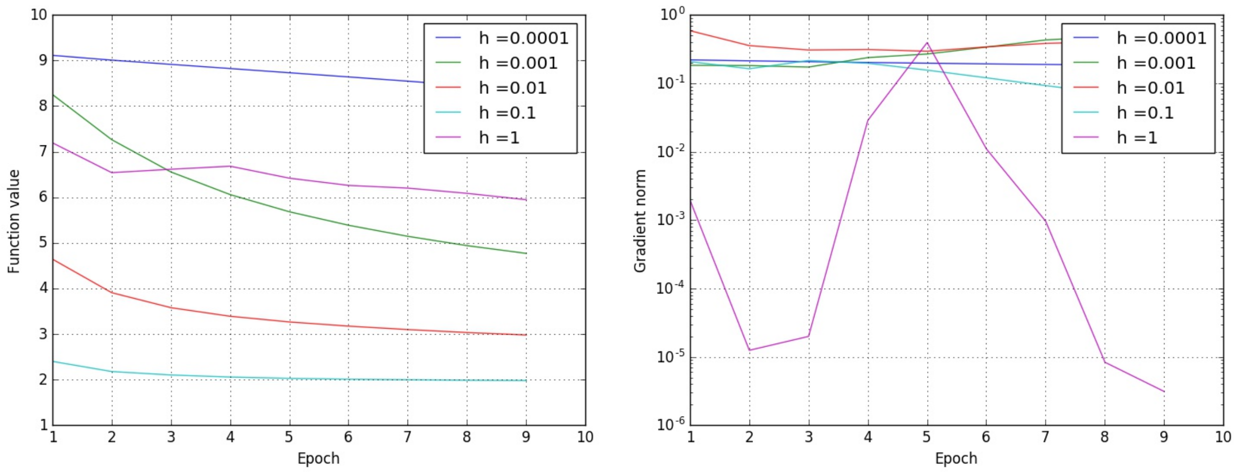
Модель: - автокодировщик
Датасет: - digits
Параметры модели: - $s = 3; d_1 = d_3 = 64; d_2 = 2; \sigma_2(x) = x; \sigma_3(x) = sigm(x)$

Результат работы метода SGD на данных digits (слева никакой оптимизации, справа работа SGD)



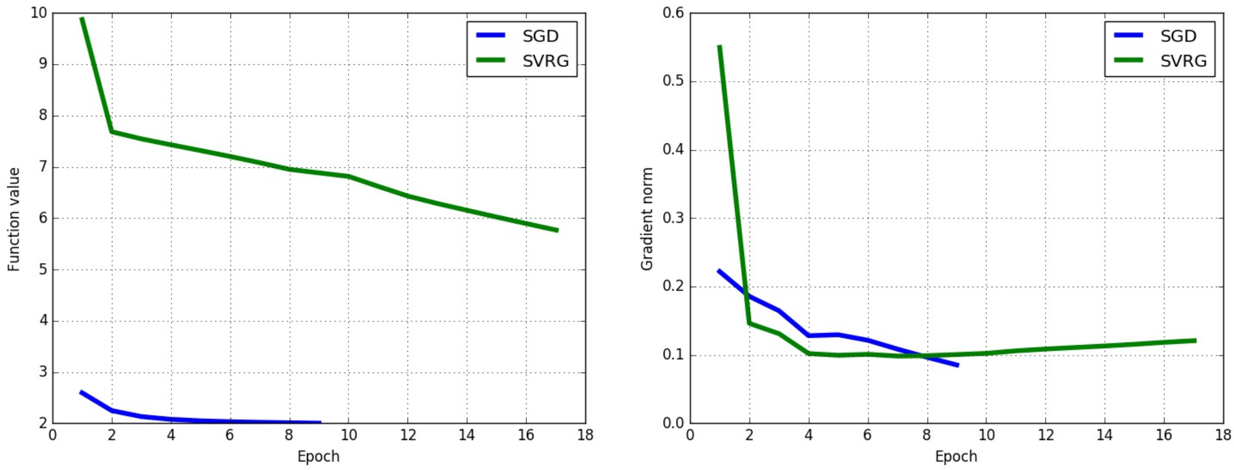


Работа метода SGD при различных длинах шага



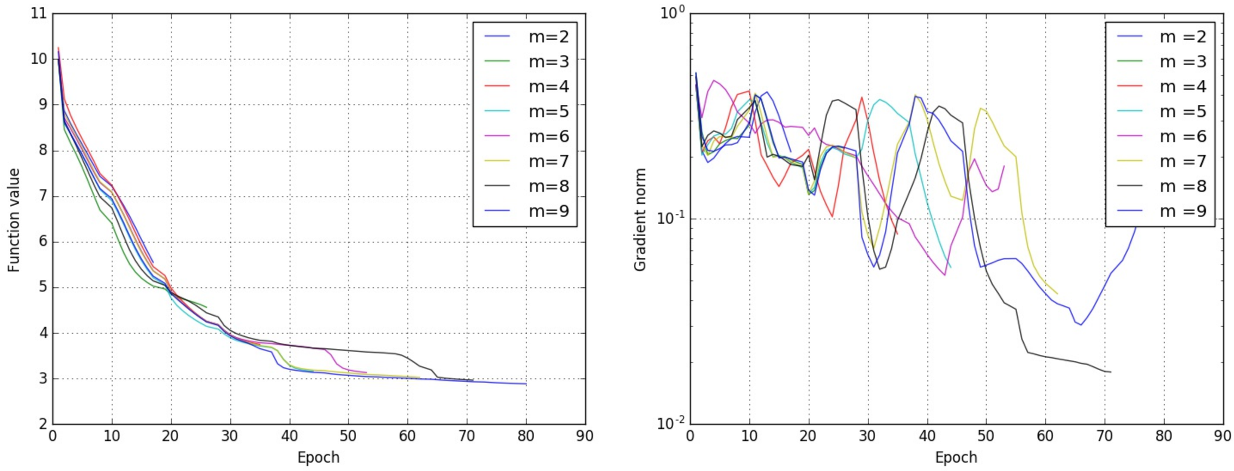
В этом примере исходная задача не является выпуклой и оснований для корректной работы алгоритма нет. Это мы и наблюдаем на графиках сходимости, хотя сходимость по значению функции похоже наблюдается, но она довольно медленная. Сходимости по градиенту совсем не наблюдается, при длине шага $h = 1$, метод ведет себя нестабильно и в итоге видимо попадает в локальный экстремум, при остальных длинах шага практически стоит на месте. Только шаг $h = 0.1$ показывает приемлимое поведение – убывание по норме градиента и минимальные значения функционала среди других вариантов шага.

Сравнение методов SGD с лучшей длиной шага и SVRG



В этой задаче SVRG состандартными параметрами показал себя довольно плохо, но зато SGD показывает довольно хороший результат, хотя задача и не является выпуклой.

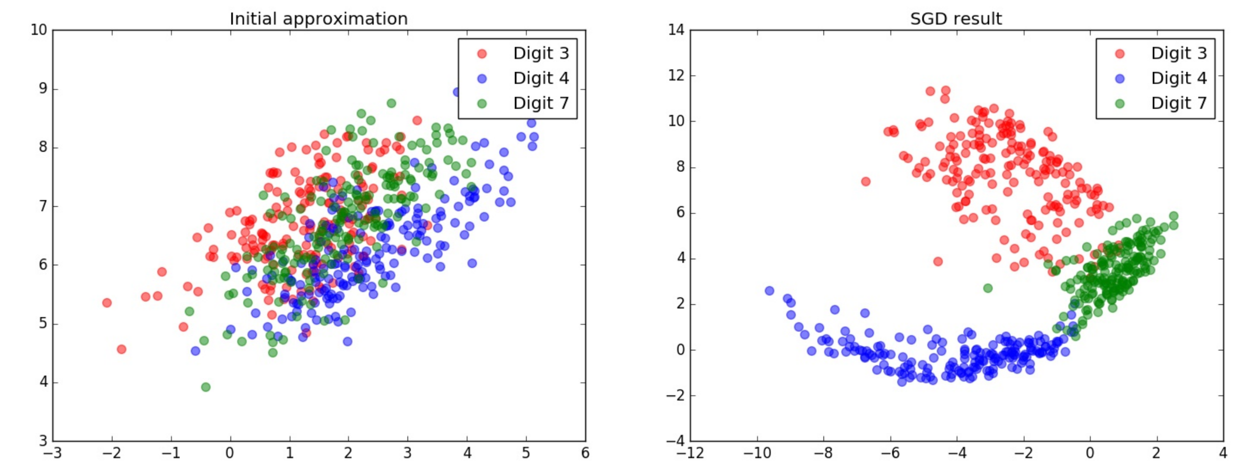
Работа метода SVRG в зависимости от длины внутреннего цикла



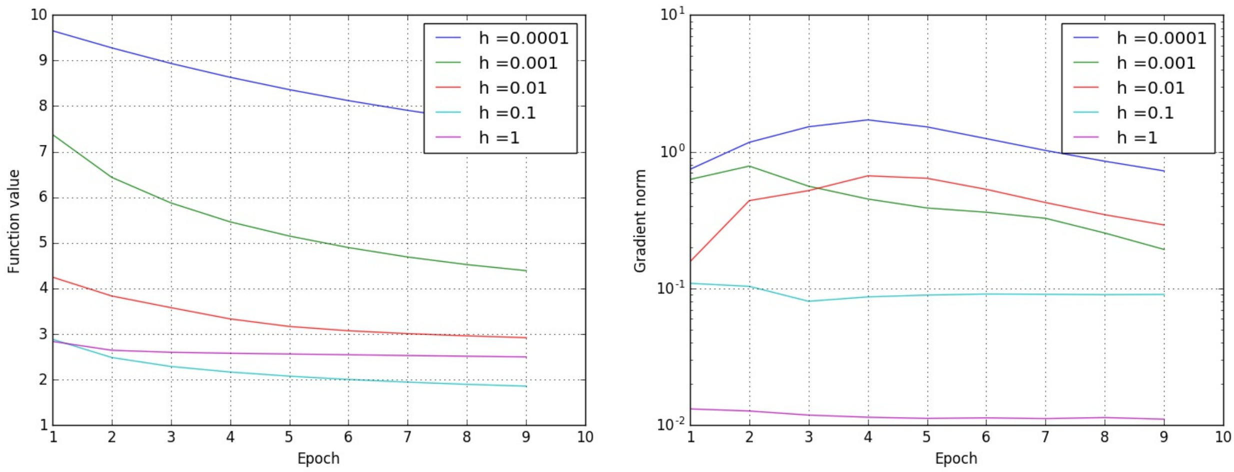
В целом графики сходимости по значению функции убывающие, как и в прошлых экспериментах, чем больше число итераций внутреннего цикла, тем меньше найденное значение. Но графики нормы градиента совсем немонотонные, метод ведет себя нестабильно, хотя общий тренд убывания нормы градиента просматривается для всех значений числа внутренних итераций.

4 Тестирование на датасете digits, архитектура II

Модель: - автокодировщик
Датасет: - digits
Параметры модели: - $s = 5; d_1 = d_5 = 64; d_2 = d_4 = 30; d_3 = 2; \sigma 3(x) = x; \sigma 2(x) = \sigma 4(x) = \sigma 5(x) = sigm(x)$
Результат работы метода SGD на данных digits (слева никакой оптимизации, справа работа SGD)

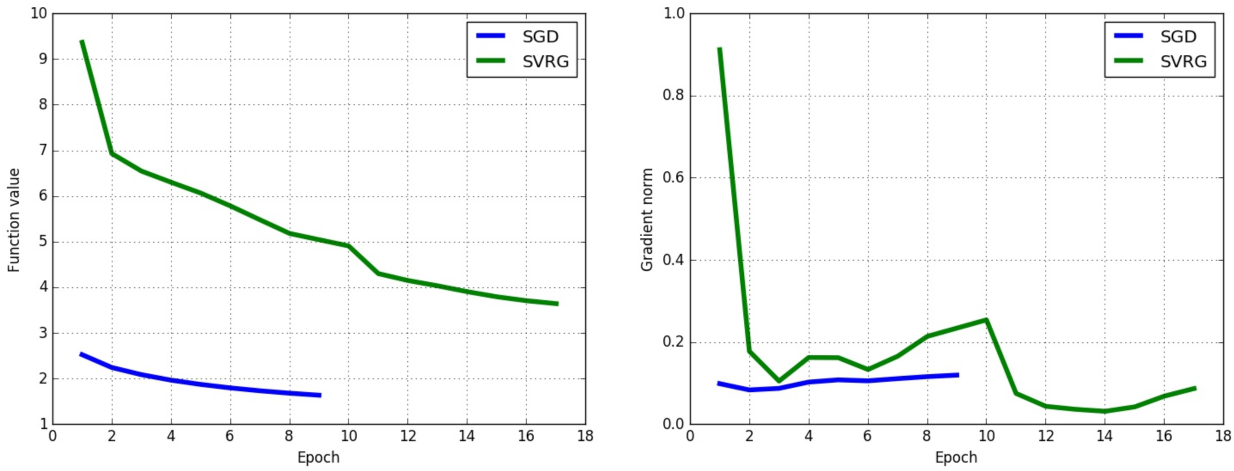


Работа метода SGD при различных длинах шага



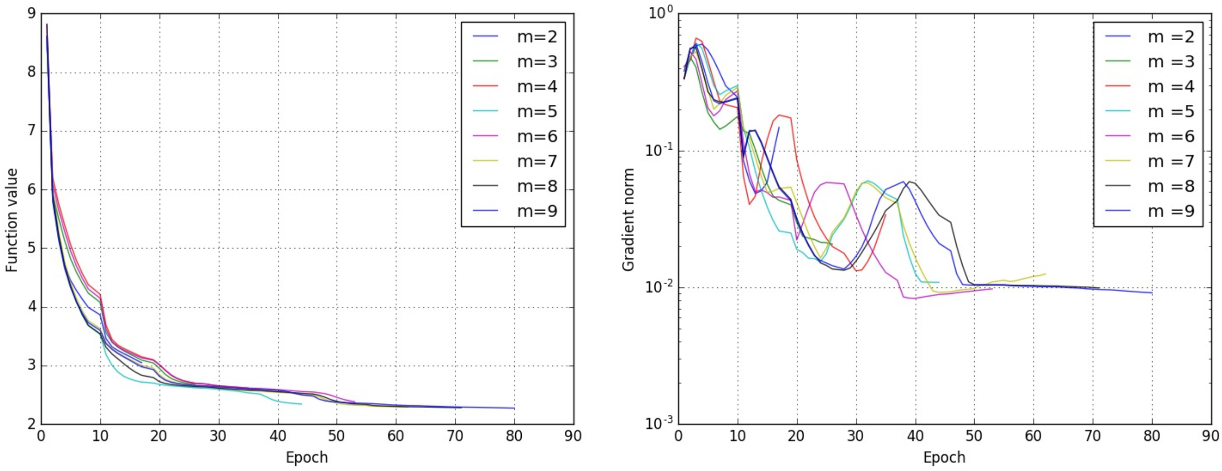
На этой задаче метод SGD работает довольно плохо, сходимости по норме градента не получилось ни для одной длины шага. Метод либо "топтался на месте"как при $h = 0.1$ и $h = 1$ либо вовсе не шел в направлении спуска. При $h = 0.1$ метод показал лучший из всех остальных длин шага результат.

Сравнение методов SGD с лучшей длиной шага и SVRG



Оба метода показали монотонное убывание значения функции, но по значению нормы градиента монотонности нет и оба метода достигли довольно малую точность. Метод SGD этом случае отработал лучше чем SVRG о стандартными параметрами.

Работа метода SVRG в зависимости от длины внутреннего цикла



Графики очень похожи на аналогичные из предыдущего примера. В целом метод оптимизирует функцию, сходимость по норме градиента немонотонная, есть убывающий тренд по норме градиента. Чем больше количество итераций внутреннего шага тем лучше результат оптимизации для SVRG.

5 Выводы

1. В случае выпуклой задачи оба метода показывают хорошие результаты, но сказать какой метод работает лучше - нельзя. В первой задаче лучше себя показал SGD, во второй - SVRG
2. В случае невыпуклой задачи SGD показал лучшие результаты