

Содержание

1 Исследовательская часть

1

1 Исследовательская часть

Задача. 1) Реализуйте Momentum и используйте его для обучения логистической регрессии. Приведите графики обучения при различных значениях гиперпараметров. Какие наблюдения и выводы можно сделать?

$$\text{momentum} = \gamma * \text{momentum} + \alpha * \text{grad}$$
$$\text{weight} = \text{weight} - \text{momentum}$$

рис 1. $\gamma = 0.9$. Особых преимуществ нет. Точность 0.96 достигается на 60000 эпохе, затем увеличивается очень медленно даже при обновлении learning rate.

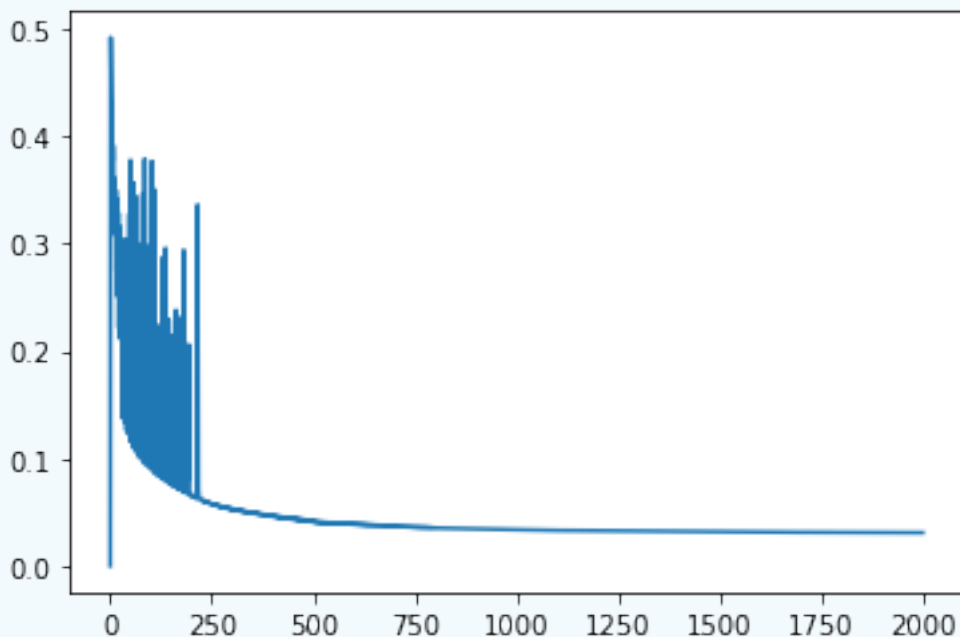


рис 2. $\gamma = 0.5$. Точность быстро выровнялась на 0.95 на 17000 эпохе обучения, 0.96 на 26000 эпохе, 0.97 не достигается даже на 100000 эпохе.

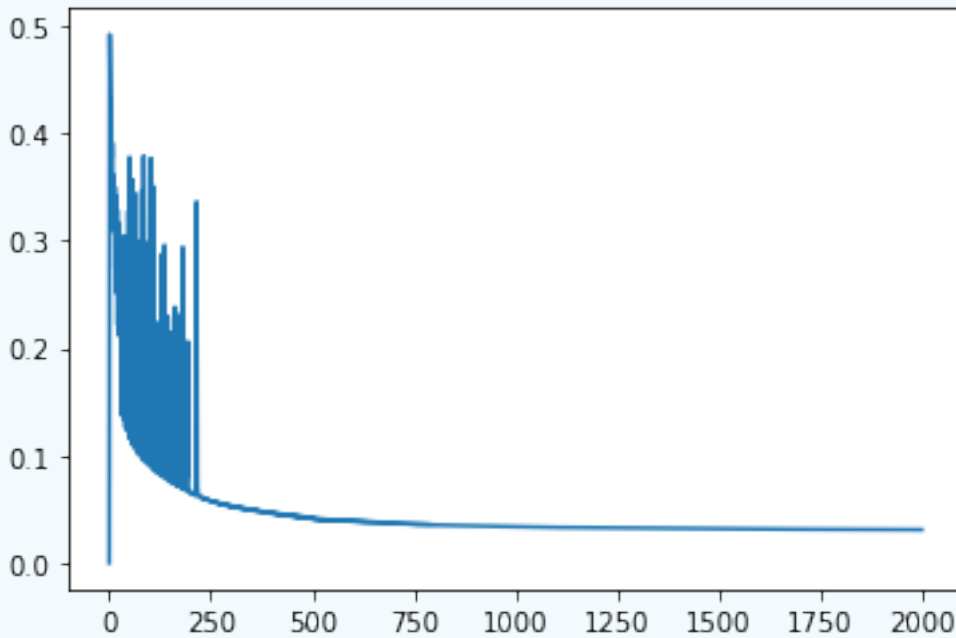
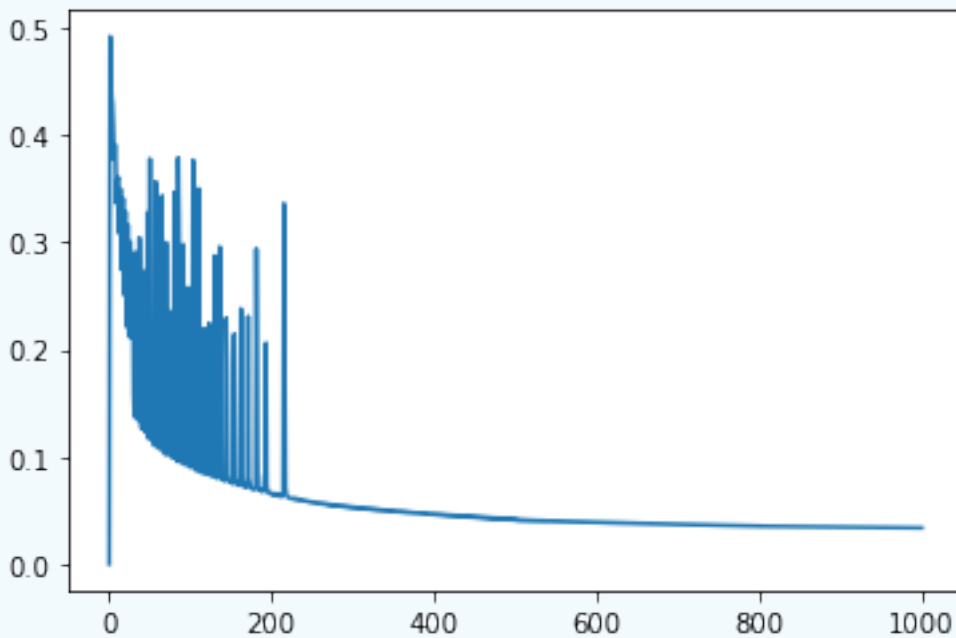


рис 3. $\gamma = 0.1$. Точность также быстро выровнялась на 0.95 на 17000 эпохе обучения, достигла 0.96 на 26000 эпохе, 0.97 не достигается даже на 100000 эпохе.



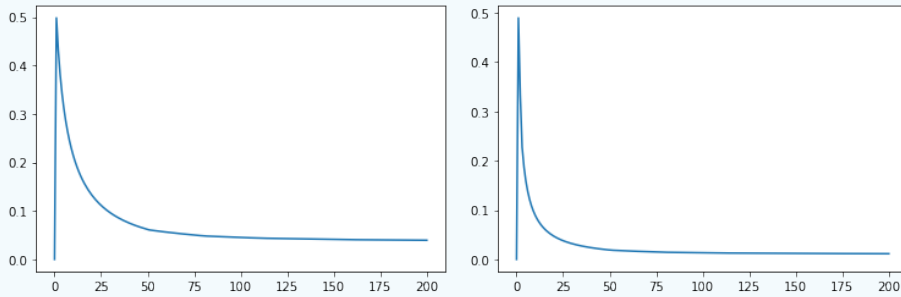
Задача. В качестве компонент bag-of-words векторов можно использовать разные признаки: абсолютные частоты слов, относительные частоты, бинарные признаки (входит слово в пример или нет). Каждый признак можно прологорифмировать, привести к диапазону $[0,1]$ или стандартизовать (вычесть среднее по обучающей выборке значение данного признака и поделить на стандартное отклонение). Попробуйте различные варианты и опишите результаты.

- 1) При использовании относительной частоты вместо абсолютной особых

преимуществ обнаружено не было. Немного увеличилось время составления матрицы из-за дополнительных подсчетов.

2) При использовании бинарных признаков увеличилось время составления матрицы из-за дополнительных подсчетов. Точность на тестовой выборке не изменилась, скорость повышения точности снизилась.

рис. 1 - бинарные признаки; рис. 2 - абсолютные частоты.



3) При логарифмировании также увеличилось время составления матрицы из-за дополнительных подсчетов. В комбинации с увеличением learning rate (и его более частым делением) точность увеличилась и достигается быстрее.

