

Домашнее задание по курсу «Машинное обучение»: SGD

Колесов Алексей & Терлиженко Ирина

8 февраля 2021 г.

1 Задания

Скачайте датасет «Ирисы Фишера» (<http://archive.ics.uci.edu/ml/datasets/Iris>). В этих данных описаны три вида ирисов. Случайным образом разбейте датасет в соотношении 9 : 1. Первую часть надо будет использовать в качестве обучающей выборки. Вторую — в качестве тестовой (для оценки качества полученной гипотезы). В каждом задании вам необходимо будет построить три модели, по одной модели для каждого вида ирисов. Соответствующая модель должна решать задачу классификации на два класса — принадлежит данный цветок рассматриваемому виду или нет.

1. **(+2 балла)** обучите модель логистической регрессии с регуляризацией Тихонова с помощью SGD-алгоритма. Необходимые параметры подберите с помощью алгоритма k-fold.
2. **(+2 балла)** обучите модель логистической регрессии с регуляризацией Тихонова с помощью алгоритма batch gradient descent. Этот алгоритм выбирает в качестве направления на каждом шаге $-\nabla L_S(w^{(t)})$, то есть эмпирический риск по всем объектам. Необходимые параметры подберите с помощью алгоритма k-fold.
3. **(+2 балла)** сравните два полученных решения. Что вы можете сказать о плюсах и минусах каждого из подходов?