

# Bayesian Methods Research Group

## Отборочное задание [deepbayes.ru](https://bayesgroup.github.io/deepbayes-school/2017/task/description/)

Хорошее знание математики очень важно в современном машинном обучении. В этом задании Вам предстоит реализовать метод машинного обучения, применив знания матричного дифференцирования, методов оптимизации и программирования с использованием векторных операций.

Вы будете решать задачу бинарной классификации: для каждого объекта нужно предсказать, к какому из двух классов он более вероятно принадлежит. Наиболее популярный линейный метод решения этой задачи - логистическая регрессия. В этой модели ответ классификатора на объекте определяется знаком скалярного произведения вектора признаков объекта на вектор весов признаков, а вектор весов настраивается с помощью метода максимального правдоподобия. Чтобы избежать переобучения, к логарифму правдоподобия часто добавляют регуляризатор, штрафующий большие значения весов. Вам предстоит реализовать логистическую регрессию, в которой оптимизируется следующая функция:

$$L(\mathbf{w}, \mathbf{X}, \mathbf{y}) = \sum_{i=1}^N \log(1 + \exp(-y_i \mathbf{x}_i^\top \mathbf{w})) + \lambda_1 \cdot \sum_{j=1}^M |w_j| + \lambda_2 \cdot \sum_{j=1}^M w_j^2 \rightarrow \min_{\mathbf{w}}$$
$$\mathbf{X} \in \mathbb{R}^{N \times M}, \mathbf{x}_i \in \mathbb{R}^M, \mathbf{w} \in \mathbb{R}^M, \mathbf{y} \in \{-1, 1\}^N$$

Здесь  $N$  число объектов в обучающей выборке,  $M$  число признаков описывающих объект,  $\mathbf{w}$  вектор весов признаков,  $\mathbf{X}$  матрица объекты-признаки,  $\mathbf{y}$  вектор ответов на объектах обучающей выборки.

Используйте файл `task.py` как заготовку решения. Вам необходимо написать тело следующих функций:

- Вычисление функции потерь `lossf(w, X, y, l1, l2)`
- Вычисление градиента функции потерь `gradf(w, X, y, l1, l2)`
- Обучение модели методом субградиентного спуска `LR.fit(self, X, y)`
- Предсказание классов для объектов `LR.predict(self, X)`
- Предсказание вероятностей `LR.predict_proba(self, X)`

Описание типов данных и размерностей аргументов приведено в заготовке решения. Язык для выполнения задания: *python 2.7*.

В качестве решения задания отправьте код, который содержит только функции `lossf`, `gradf` и класс `LR`. Не разрешается менять прототипы функций, подключать дополнительные модули, кроме: `numpy`, `scipy`, `sklearn`. Решения, не соответствующие выданному прототипу, будут автоматически засчитаны как неверные. Код функций `lossf` и `gradf` должен быть написан с использованием векторных вычислений `numpy` и не должен содержать циклов.

Мы предоставляем функцию `test_work`, которая проверяет интерфейсы, но не проверяет, что логистическая регрессия реализована верно. Правильность реализации логистической регрессии будет проверена после отправки решения.

 bayesgroup  
bayesml@gmail.com