

Задание 2. Рекуррентные нейронные сети

Курс: Глубинное обучение, осень 2016

Начало выполнения задания: 11 ноября.

Срок сдачи: **27 ноября, 23:59.**

Среда для выполнения задания: Python 2.

1 Описание

Данное домашнее задание состоит из двух основных задач:

- обучение генеративной модели текста,
- обучение классификатора последовательностей на основе рекуррентной сети.

В каждой задаче есть некоторая основная часть, в которой обучается простейшая рекуррентная модель для данной задачи. Реализация этой части достаточно подробно пошагово описывается в приложенных jupyter-notebook-ax.

Далее предлагается несколько дополнительных направлений, в которых эту модель можно улучшать. Вы можете пробовать любые из этих направлений или какие-то свои. По выбранным дополнительным направлениям нужно провести небольшое, но как можно более полное исследование: запрограммировать все необходимое, провести эксперименты и **проанализировать результаты**.

В обязательную часть задания входят:

1. основные части для обеих задач,
2. два дополнительных исследования на Ваш выбор. Исследования не должны повторяться (нельзя брать одинаковые пункты из обеих задач, например, визуализацию или обучение более сложной модели + контроль переобучения).

За выполнения большего числа дополнительных исследований будут выставляться бонусные баллы. Здесь, опять же, выполнение одинаковых пунктов будет поощряться меньше.

Большая просьба: при сдаче работы аккуратно прописывайте в ноутбуках что именно вы делаете после основной части, чтобы проверяющему было понятно что к чему. Если в основной части Вы используете другой набор данных, то четко опишите его (объем данных + постановка задачи + вставьте небольшой кусок данных в ноутбук). Присылать данные с заданием не нужно!

Генерация текста. В данном задании необходимо обучить генеративную модель текста. К заданию приложены текстовые файлы с кодексами законов и Вам нужно сгенерировать свою новую конституцию (идея взята из домашнего задания курса «Обучение представлений и глубокое обучение» в ШАДе), однако вы можете обучаться на любых других текстовых данных, которые вам нравятся.

Классификация последовательностей. В данном задании необходимо обучить бинарный классификатор, который отличает мужские имена от женских. Если у вас есть на примете другая задача классификации последовательностей, то вы можете решать ее. Например, можно отличать тексты на разных языках или стихи от прозы.

2 Оформление задания

Выполненное задание следует отправить письмом по адресу bayesml@gmail.com с заголовком письма

«[ВМК ГО16] Задание 2, Фамилия Имя».

Убедительная просьба присылать выполненное задание только один раз с окончательным вариантом.

Присланный вариант задания должен содержать в себе два ноутбука со всеми исследованиями. Ноутбуки должны называться *surname_generation* и *surname_classification*, где surname — Ваша фамилия на английском языке с маленькой буквы.

Убедительная просьба не изменять прототипы функций в основной части задания. Если Вы используете другие данные в задании, то присылать их не нужно, достаточно описать в ноутбуке. Если данные понадобятся для проверки, мы напишем Вам об этом дополнительно.