

# Машинное обучение: задание 4

Составитель: Виктор Кантор

28 марта 2017 г.

## Организационные вопросы

### Дедлайн

На выполнение задания дается две недели, последний срок сдачи — 11 апреля 23:59.

### Сдача задания

1. Пришлите на почту `ml.course.mipt@gmail.com` письмо с решениями заданий в `ipynb//py//pdf` файлах (где что более уместно). Тему письма укажите в формате «ML2017 <номер\_группы> <фамилия> <имя>, Задание 4 (Linear models)», например: «ML2017 492 Страуструп Бьярн, Задание 4 (Linear models)»
2. Если есть принципиальное желание оформлять теоретические задачи в `ipynb` в Markdown-ячейках, это не запрещается. Также не запрещается решать теоретические задачи на бумаге, оформлять их аккуратно и разборчиво, и присылать pdf со сканами, вместо того, чтобы набирать в L<sup>A</sup>T<sub>E</sub>X.

## Контрольные вопросы

Ниже приводится список вопросов, с ответами на которые может быть полезно разобраться для понимания темы.

1. Как выглядит решающее правило в линейной классификации? А зависимость, которой мы приближаем ответы в линейной регрессии?
2. Что такое функции потерь в задачах классификации и регрессии? Зачем они нужны?
3. Что такое регуляризаторы? Какими они бывают в задачах линейной классификации и регрессии? Зачем они нужны?
4. Как в общем виде выглядит оптимизационная задача в линейной классификации или линейной регрессии?
5. Как работает настройка весов в линейной модели с помощью SGD (Stochastic Gradient Descent)? Как выглядит правило обновления весов?
6. Учитывается ли коэффициент сдвига  $w_0$  в регуляризаторе? Почему?
7. Почему линейные модели рекомендуется применять к выборке с нормированными значениями признаков?
8. Как выглядит оптимизационная задача в логистической регрессии? А в SVM?
9. Выпишите и докажите формулу для весов в линейной регрессии (с квадратичной функцией потерь). То же самое для гребневой регрессии.
10. Выпишите SGD для логистической регрессии с  $\ell_2$ -регуляризацией и для SVM с линейным ядром.
11. В чем заключается идея ядер в SVM?
12. Какие преимущества и недостатки есть у линейных моделей?
13. Что такое Precision и Recall? А ROC-AUC? Когда он применяется?

# 1 Контест на линейные модели и генерацию признаков

30% баллов за задание

Описание контеста - на странице соревнования, присоединиться можно по ссылке:  
[https://kaggle.com/join/diht\\_linear\\_models\\_1](https://kaggle.com/join/diht_linear_models_1) (ссылка заработает после прохождения контестом модерации на kaggle inclass)

В названии команды обязательно указывайте номер своей группы, имя и фамилию.

Уточненная разбалловка по заданию появится позже.

# 2 Контест на линейные модели и сведение задачи к стандартной

30% баллов за задание

Описание контеста - на странице соревнования, присоединиться можно по ссылке:  
[https://kaggle.com/join/diht\\_linear\\_models\\_2](https://kaggle.com/join/diht_linear_models_2) (ссылка заработает после прохождения контестом модерации на kaggle inclass)

В названии команды обязательно указывайте номер своей группы, имя и фамилию.

Уточненная разбалловка по заданию появится позже.

# 3 Теоретические задачи

40% баллов за задание

## 3.1 Знакомство с линейным классификатором

1. Как выглядит бинарный линейный классификатор? (Формула для отображения из множества объектов в множество классов.)
2. Что такое отступ алгоритма на объекте? Какие выводы можно сделать из знака отступа?
3. Как классификаторы вида  $a(x) = \text{sign}(\langle w, x \rangle - w_0)$  сводят к классификаторам вида  $a(x) = \text{sign}(\langle w, x \rangle)$ ?
4. Как выглядит запись функционала эмпирического риска через отступы? Какое значение он должен принимать для «наилучшего» алгоритма классификации?
5. Если в функционале эмпирического риска (риск с пороговой функцией потерь) всюду написаны строгие неравенства ( $M_i < 0$ ) можете ли вы сразу придумать параметр  $w$  для алгоритма классификации  $a(x) = \text{sign}(\langle w, x \rangle)$ , минимизирующий такой функционал?
6. Запишите функционал аппроксимированного эмпирического риска, если выбрана функция потерь  $L(M)$ .
7. Что такое функция потерь, зачем она нужна? Как обычно выглядит ее график?
8. Приведите пример негладкой функции потерь.
9. Что такое регуляризация? Какие регуляризаторы вы знаете?
10. Как связаны переобучение и обобщающая способность алгоритма? Как влияет регуляризация на обобщающую способность?
11. Как связаны острые минимумы функционала аппроксимированного эмпирического риска с проблемой переобучения?
12. Что делает регуляризация с аппроксимированным риском как функцией параметров алгоритма?

13. Для какого алгоритма классификации функционал аппроксимированного риска будет принимать большее значение на обучающей выборке: для построенного с регуляризацией или без нее? Почему?
14. Для какого алгоритма классификации функционал риска будет принимать большее значение на тестовой выборке: для построенного с оправдывающей себя регуляризацией или вообще без нее? Почему?
15. Что представляют собой метрики качества Ассурасу, Precision и Recall?
16. Что такое метрика качества AUC и ROC-кривая?
17. Как построить ROC-кривую (нужен алгоритм), если например, у вас есть правильные ответы к домашнему заданию про фамилии и ваши прогнозы?

### 3.2 Вероятностный смысл регуляризаторов

Покажите, что регуляризатор в задаче линейной классификации имеет вероятностный смысл априорного распределения параметров моделей. Какие распределения задают  $l_1$ -регуляризатор и  $l_2$ -регуляризатор?

### 3.3 SVM и максимизация разделяющей полосы

Покажите, как получается условная оптимизационная задача, решаемая в SVM из соображений максимизации разделяющей полосы между классами. Можно отталкиваться от линейно-разделимого случая, но итоговое выражение должно быть для общего.

Как эта задача сводится к безусловной задаче оптимизации?

### 3.4 Kernel trick

Придумайте ядро, которое позволит линейному классификатору с помощью Kernel Trick построить в исходном пространстве признаков разделяющую поверхность  $x_1^2 + 2x_2^2 = 3$ . Какой будет размерность спрямляющего пространства?

### 3.5 $\ell_1$ -регуляризация

Покажите с помощью теоремы Куна-Таккера, что ограничение  $\ell_1$ -нормы вектора весов числом и добавление штрафа с его  $\ell_1$ -нормой приводят к построению одного и того же алгоритма. Можно считать, что регуляризатор добавляется по существу, т.е. меняет итоговый ответ по сравнению с оптимизационной задачей без регуляризатора.

### 3.6 Повторение: метрики качества

1. Что представляют собой метрики качества Ассурасу, Precision и Recall?
2. Что такое метрика качества AUC и ROC-кривая?
3. Как построить ROC-кривую (нужен алгоритм), если например, у вас есть правильные ответы к домашнему заданию про фамилии и ваши прогнозы?