Министерство образования и науки Российской Федерации

Калужский филиал

федерального государственного бюджетного образовательного

учреждения высшего образования

**«Московский государственный технический университет**

**имени Н.Э. Баумана**

**(национальный исследовательский университет)»**

(КФ МГТУ им. Н.Э. Баумана)

М.Г. Семененко

(к.ф.-м.н. доцент)

**ЛАБОРАТОРНАЯ РАБОТА № 1**

**по курсу «Методы машинного обучения»**

## ЛИНЕЙНЫЕ КЛАССИФИКАТОРЫ и их свойства

Калуга

2018

**Теоретические основы.**

Машинное обучение (machine learning) – это область научного знания, имеющая дело с алгоритмами, "способными обучаться". Необходимость использования методов машинного обучения объясняется тем, что для многих сложных – "интеллектуальных" – задач (например, распознавание рукописного текста, речи и т. п.) очень сложно (или даже невозможно) разработать "явный" алгоритм их решения, однако часто можно научить компьютер обучиться решению этих задач.

Задача обучения по прецедентам заключается в том, чтобы по выборке  восстановить зависимость , то есть построить решающую функцию (decisionfunction): , которая приближала бы целевую функцию , причём не только на объектах обучающей выборки, но и на всём множестве .

Решающая функция должна допускать эффективную компьютерную реализацию; по этой причине будем называть её алгоритмом.

Метод обучения (learning algorithm) — это отображение *µ*: (*X* × *Y* )*ℓ* → *A*, которое произвольной конечной выборке ставит в соответствие некоторый алгоритм *a* ∈ *A*. Говорят также, что метод *µ* строит алгоритм *a* по выборке *Xℓ*. Метод обучения должен допускать эффективную программную реализацию.

На этапе обучения метод *µ* по выборке *Xℓ* строит алгоритм *a* = *µ*(*Xℓ*).

На этапе применения алгоритм *a* для новых объектов *x* выдаёт ответы *y* = *a*(*x*).

Этап обучения наиболее сложен. Как правило, он сводится к поиску параметров модели, доставляющих оптимальное значение заданному функционалу качества.

Функция потерь (loss function) — это неотрицательная функция L (*a,x*), характеризующая величину ошибки алгоритма *a* на объекте *x*. Если L (*a,x*) = 0, то ответ *a*(*x*) называется корректным.

Функционал качества алгоритма *a* на выборке *Xℓ*:

Изображение выглядит как черный, Шрифт, рукописный текст, доска

Содержимое, созданное искусственным интеллектом, может быть неверным.*.* (1.2)

Функционал *Q* называют также функционалом средних потерь или эмпирическим риском, так как он вычисляется по эмпирическим данным .

Функция потерь, принимающая только значения 0 и 1, называется бинарной. В этом случае L (*a,x*) = 1 означает, что алгоритм *a* допускает ошибку на объекте *x*, а функционал *Q* называется частотой ошибок алгоритма *a* на выборке *Xℓ*.

Наиболее часто используются следующие функции потерь, при *Y* ⊆ R:

L (*a,x*) = [*a*(*x*) =6 *y*∗(*x*)] — индикатор ошибки, обычно применяется в задачах классификации;

L (*a,x*) = |*a*(*x*) − *y*∗(*x*)| — отклонение от правильного ответа; функционал *Q* называется средней ошибкой алгоритма *a* на выборке *Xℓ*;

L (*a,x*) = (*a*(*x*) − *y*∗(*x*))2 — квадратичная функция потерь; функционал *Q* называется средней квадратичной ошибкой алгоритма *a* на выборке *Xℓ*; обычно применяется в задачах регрессии.

**Решение задачи регрессии в электронных таблицах Excel**

1. Исходные данные:

Изображение выглядит как текст, число, Шрифт, чек

Содержимое, созданное искусственным интеллектом, может быть неверным.

Найдем уравнение линейной и квадратичной регрессии:

1. Вводим массив исходных данных.
2. Строим точечную диаграмму данных.
3. Выделяем точки на диаграмме, щелкаем правой кнопкой мыши и выбираем **Добавить линию тренда**. Отмечаем пункты **Показывать уравнение регрессии на диаграмме**

Результаты моделирования представим в виде таблицы:

Табл. 1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X | Y | Yлин | Yкв | ⎢Y − Yлин⎢ | ⎢Y − Yкв⎢ |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| … | … | … | … | … | … |  |
| Итого |  |  |  | Σ | Σ |  |

Используя функцию СУММКВ, вычислить среднюю ошибку и квадратичную функцию потерь.

1. Провести приближение функции Рунге полиномами 5-й и 6-й степени. Массив Х задать на отрезке [-2; 2] для 15 точек (включая концы отрезка). Представить результаты в виде графиков и таблицы 1.
2. Вычислить среднюю ошибку и квадратичную функцию потерь в каждом случае.

### 