**Варианты лабораторной работы №1**

**по курсу «Введение в машинное обучение»**

**Файл с данными**

data\_vN-NN.csv, N-NN – номер варианта

**Шаблон отчета**

<https://docs.google.com/presentation/d/1SaG1JYkH_X5eFrq6VleHDd757jmKna8wxb3YyGyK7h0/edit?usp=sharing>

**Отчет**

Файл otchet\_vN-NN\_GroupFIO.pdf и **исходные коды**, написанные для выполнения заданий, высылать на [angultiaev@gmail.com](mailto:angultiaev@gmail.com)

**Указания по оформлению графиков**

<http://datalearning.ru/study/Courses/methodic/lections/graphs.pdf>

К каждому графику в отчете должны быть даны все необходимые пояснения для его понимания (название, подписи осей, легенда, условия эксперимента, параметры обработки и пр.).

**Рекомендуемые средства программной реализации**

Python, MATLAB

**Литература**

1. Alpaydin, E. (2014). *Introduction to machine learning*. MIT press.

2. Shalev-Shwartz, S., & Ben-David, S. (2014). *Understanding machine learning: From theory to algorithms*. Cambridge university press.

3. Raschka, S. (2015). *Python machine learning*. Packt Publishing Ltd.

4. Fox J. (2015). Applied regression analysis and generalized linear models. Sage Publications.

5. Sen, Ashish, and Muni Srivastava. *Regression analysis: theory, methods, and applications*. Springer Science & Business Media, 1997.

**Вариант 1-09**

**Nonlinear least-squares**

**Задания**

1. Построить полиномиальные регрессионные модели при *m*=1,2,3, используя метод OLS, рассчитать коэффициенты детерминации моделей на обучающей и тестовой выборках. На диаграмме рассеяния исходных данных вывести рассчитанные функции регрессии.

2. Обучить нелинейную экспоненциальную модель, рассчитать коэффициент детерминации модели на обучающей и тестовой выборках, сравнить результаты с полиномиальными моделями.

3. Построить зависимость значения MSE от параметров нелинейной модели (визуализировать в виде поверхности и контурной диаграммы). Определить визуально оптимальные значения параметров и сравнить их со значениями, найденными в результате обучения.

4. Применить логарифмическое преобразование результирующей переменной и построить простейшую линейную регрессионную модель. Рассчитать коэффициент детерминации модели на обучающей и тестовой выборках, визуализировать полученную функцию регрессии, сравнить с результатами пп. 1, 2.

5. Построить гистограммы распределения остатков моделей, обученных в пп. 1, 2, 4 на обучающей и тестовой выборках, проверить распределения остатков на нормальность, используя критерий «хи-квадрат» (привести значения статистики критерия, *p-value*, статистическое решение). Оценить нормальность остатков визуально по Q-Q диаграммам.

6. Сделать вывод по результатам сравнения обученных линейных и нелинейной регрессионных моделей.

**Указания**

Класс полиномиальных регрессионных моделей: .

Класс экспоненциальных регрессионных моделей:  .

Кросс-валидация: Holdout (70/30).

Для обучения нелинейной регрессионной использовать один из методов оптимизации (простой градиентный метод, метод Левенберга-Маркардта или др.).

Выводить графики MSE в логарифмическом масштабе.

**Литература**

1. Sen, Ashish, and Muni Srivastava. *Regression analysis: theory, methods, and applications*. Springer Science & Business Media, 1997.

2. Seber, George AF, and Christopher John Wild. "Nonlinear regression." *New Jersey: John Wiley & Sons* 62 (2003): 63.