Лабораторная работа #3 часть 2

**Применение логистической регрессии для решения задачи бинарной классификации**

**Часть 1.** В данной части работы мы познакомимся с датасетом Titanic. В дальнейшем данный датасет мы будем использовать в качестве тестового для сравнения различных алгоритмов классификации, поэтому **эту часть работы необходимо выполнить в отдельном блокноте**, который мы будем дополнять.

Данный набор данных содержит информацию о пассажирах Титаника. Набор содержит признак Survived для каждого пассажира, обозначающий, выжил данный пассажир или нет (0 для умерших, 1 для выживших), который и будет нашей целью.

Каждая строчка наборов данных содержит следующие поля:

* Pclass — класс пассажира (1 — высший, 2 — средний, 3 — низший);
* Name — имя;
* Sex — пол;
* Age — возраст;
* SibSp — количество братьев, сестер, сводных братьев, сводных сестер, супругов на борту титаника;
* Parch — количество родителей, детей (в том числе приемных) на борту титаника;
* Ticket — номер билета;
* Fare — плата за проезд;
* Cabin — каюта;
* Embarked — порт посадки (C — Шербур; Q — Квинстаун; S — Саутгемптон).

Задание 1. Выполнить предобработку данных с помощью Pandas

* 1. Удалить все строки данных содержащие пропуски.
  2. Удалить столбцы содержащие не числовые значения, за исключением столбцов Sex и Embarked
  3. Перекодировать данные в числовой вид в столбцах Sex и Embarked. Sex – 0 и 1, Embarked – 1,2,3
  4. Удалить столбец PassengerId как не несущий полезной информации.
  5. Вычислить процент потерянных данных после выполнения пп 1.1 – 1.2

Задание 2. Машинное обучение

2.1. Разбить полученные данные на обучающую и тестовые выборки

2.2. Решить задачу классификации методом логистической регрессии

2.3. Вывести точность полученной модели

2.4. Оценить влияние признака Embarked на точность модели

Часть 2. Многоклассовая логистическая регрессия

Задание: Для рассмотренного в предыдущей работе набора данных Iris, решить задачу многоклассовой логистической регрессии. Отрисовать полученное решение на плоскости 'petal length - petal width'. Выполнить в отдельном блокноте.

Отчет по работе должен представлять zip архив содержащий 2 блокнота с частями работы.