**Варианты лабораторной работы №2**

**по курсу «Анализ данных и 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.

**Вариант 2-09**

**Normal Bayes classifier**

**Задания**

1. Оценить ковариационные матрицы обучающих данных для каждого класса при различных предположениях: а) ковариационные матрицы всех классов равные, скалярные; б) равные, диагональные; в) различные, скалярные; г) различные, диагональные; д) равные; е) различные. Визуализировать матрицы для каждого случая.

2. Рассчитать классификационные очки обученных байесовских классификаторов (при предположениях а)–е) п.1) для примеров обучающей и тестовой выборок.

3. В исходном пространстве признаков изобразить области классов (закрасить разными цветами), формируемые каждым из обученных классификаторов. Нанести на диаграммы границы классов и данные из обучающей и тестовой выборок. На отдельной диаграмме изобразить все границы классов, формируемые построенными классификаторами.

4. Рассчитать точности (accuracy) построенных классификаторов на обучающей и тестовой выборках, представить данные в графическом виде (bar charts).

5. Для каждого классификатора построить micro-averaged и macro-averaged ROC-кривые и PR-кривые на обучающей и тестовой выборках и рассчитать micro-averaged и macro-averaged ROC AUC и PR AUC на обучающей и тестовой выборках.

6. Сделать выводы о влиянии предположений о модели обучающих данных на точность байесовской классификации.

**Указания**

Кросс-валидация: Holdout (70/30).

Априорные вероятности классов считать равными.

Классификационные очки классификаторов полагать равными апостериорным вероятностям классов.

**Литература**

1. Figueiredo, M. A. (2004). Lecture notes on bayesian estimation and classification. *Instituto de Telecomunicacoes-Instituto Superior Tecnico*, *60*.

2. Michalski, R. S., Carbonell, J. G., & Mitchell, T. M. (Eds.). (2013). *Machine learning: An artificial intelligence approach*. Springer Science & Business Media.