Комментарии к лабораторным работам:
Исходя из того, что у некоторых студентов возникли вопросы по поводу того, почему условия для лабораторных оформлены не до конца, объясняю правила приема задач. Данный курс направлен на то, чтобы вы усвоили некоторое положение дел в машинном обучении: основные алгоритмы, методики и подходы. Тут нет цели обучить вас пользоваться языком Python или стандартными библиотеками для МЛ (с этим вы справитесь и сами). Поэтому при проверке лабораторных работ я сначала задаю какие-то **простые** вопросы, а только потом смотрю то, что вы написали. Вопросы эти на понимание того, что вы делаете и не представляют трудности, если вы достаточно хорошо отдаете себе отчет в своих действиях. Но, чтобы вам было полегче, я напишу некоторые вопросы, которые могу задать, а могу спросить и другие вопросы исходя из нашей с вами беседы.

0. Написать скрипт для построения ROC-кривой, PR-кривой, нахождения площади под обеими кривыми.

   Комментарий: тут я могу попросить вас вывести данные кривые для каких либо классификаторов.

1. Реализуйте метод потенциальных функций на примере датасета iris. Данный датасет лучше скачивать отсюда
   http://scikit-learn.org/stable/auto_examples/datasets/plot_iris_dataset.html

   Комментарий: продемонстрируйте мне, что ваш алгоритм работает правильно. Посмотрите основные свойства, особенности классификатора. Могу попросить вас сделать некоторую визуализацию работы данного классификатора (вывести ошибочные предсказания, еще что-то).

2. Задача на байесовский классификатор, которую я еще не оформил.

3. Возьмите датасет Mnist (рукописные цифры от 0 до 9) и используйте каждый из известных вам классификаторов, сравните качество классификации, объясните почему одни из классификаторов работают лучше или хуже. Загрузить его можно следующим образом:

   ```
   from sklearn.datasets import fetch_mldata
   mnist = fetch_mldata('MNIST original')
   ```

   Комментарий: напишите процедуру обучения правильно (кросс валидация, все дела). Я могу попросить вас реализовать какой либо использованный вами классификатор самостоятельно, либо задать вопрос на понимание принципов

работы какого-либо из них. Могу попросить реализовать один из способов сведения задачи многоклассовой классификации к бинарной на поднаборе данных.

4. *Задача на композиции алгоритмов.*
   Вам предложена задача предсказания стоимости домов в Америке (что ли). Воспользуйтесь известными библиотеками (sklearn, xgboost, lightgbm и catboost) для решения данной задачи. Попробуйте каждую из предложенных библиотек, оценить ее плюсы и минусы. Приветствуются различные методики ансамблирования (стекинг, блендинг). Данные используйте вот отсюда: https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data разбивайте на train и тест с seed=98987. Можете попробовать засабмитить свою модель на соревнования.

   Комментарий: Я могу попросить сравнить различные способы композиций алгоритмов, какие-то попрошу написать самостоятельно (ну, например, бэггинг по подпространствам). Ну и разумеется, вы должны быть в состоянии мне объяснить каждое ваше действие и базовые принципы работы композиций алгоритмов.

5. ***Конкурсная задача.***
   Для решения конкурсной задачи вам необходимо зарегистрироваться на ресурсе по соревнованиям https://www.kaggle.com.

Итоги конкурса подводятся 30.12.18

Названия ноутбуков следующего вида: группа_фамилияинициалы. Пример: 16231_GoncharenkoAI.

Комментарий: задача обязательна к решению для всех, лучшие получат небольшие бонусы к экзамену. Откровенно плохие решения будут приниматься соответственно. В данной задаче я не буду спрашивать вас алгоритмы, а скорее то, как вы работаете именно с самими данными, почему пошли так, а не иначе и тд.

Награды за призовые места внутри группы:
I место: + 1 балл к итоговой оценке и задача 4 засчитывается автоматом
II место:+ 0.5 балл к итоговой оценке и задача 4 засчитывается автоматом
III место: задача 4 засчитывается автоматом