ПСМО ФКН ВШЭ, 3 курс, 1 модуль

Задание 2. Тестирование гипотез.

Прикладная статистика в машинном обучении, осень 2019

Время выдачи задания: 14 октября.

Срок сдачи: **28 октября (понедельник), 23:59.**

Среда для выполнения практического задания – PYTHON 2.x/PYTHON 3.x.

Правила сдачи

Инструкция по отправке:

1. Решения задач следует присылать единым файлом формата `pdf`, набранным в LaTeX, либо в составе ipython-тетрадки в форматах `ipynb` и `html` (присылайте оба формата, т.к. AnyTask из-за высокой загрузки иногда не рендерит тетрадки в формате `ipynb` – а если мы не увидим ваши задачи, мы их не проверим). Отправляйте практические задачи в виде отдельных файлов (ipython-тетрадок или исходных файлов с кодом на языке `python`).

Оценивание и штрафы:

1. Максимально допустимая оценка за работу над основными задачами – 10 баллов.

2. Бонусные баллы (см. конец домашнего задания) и влияют на освобождение от задач на экзамене.

3. Дедлайн жесткий. Сдавать задание после указанного срока сдачи нельзя.

4. Задание выполняется каждым студентом индивидуально и независимо от других студентов. «Похожие» решения считаются плагиатом и все студенты (в том числе те, у кого списали) не могут получить за него больше 0 баллов, причем обнуляются и бонусные баллы. Если вы нашли решение какого-то из заданий (или его часть) в открытом источнике, необходимо указать ссылку на этот источник в отдельном блоке в конце вашей работы (скорее всего вы будете не единственным, кто это нашел, поэтому чтобы исключить подозрение в плагиате, необходима ссылка на источник).

Основные задачи

1. (2 балла) Постройте распределение плотности вероятностей для p-значения в случаях, если нулевая гипотеза верна и не верна, для следующих случаев:

   - (1 балл) одновыборочный t-тест для $X_1, \dots X_n \sim \mathcal{N}(x; \mu, 1)$, для проверки нулевой гипотезы H_0: $\mu_0 = 0$, против альтернативной гипотезы H_0: $\mu_0 > 0$. Для $\mu = 0$, $\mu = 0,5$ и $\mu = 1$. Тестирование проводите в случае, если $n = 4$.
   - (1 балл) одновыборочный t-тест для $X_1, \dots X_n \sim \exp(x; 1)$, для проверки нулевой гипотезы H_0: $\mu_0 = 0$, против альтернативной гипотезы H_0: $\mu_0 > 0$. Тестирование проводите в случае, если $n = 4, 10, 100$. Сделайте вывод о применимости метода анализа множественного тестирования Storey-Tibshirani (лекция, раздел FDR и Теорема байеса).

2. (3 балла) Вы задались целью статистически достоверно сравнить качество двух стохастических алгоритмов машинного обучения (например, алгоритмов из семейства reinforcement learning). Предположим, что качество алгоритма 1 задается (случайной) величиной X_1, а качество алгоритма 2 – величиной X_2 (распределения X_1 и X_2 неизвестны). Алгоритм 1 назовем неразличимым по качеству с алгоритмом 2, если их средние уровни качества равны: $\Delta\mu = \mu_1 - \mu_2 = \mathrm{E}X_1 - \mathrm{E}X_2 = 0$; в противном случае алгоритм 1 лучше (хуже) по качеству, чем алгоритм 2.

   Для того, чтобы сравнивать алгоритмы по качеству, воспользуемся аппаратом проверки статистических гипотез. Таким образом, для сравнения алгоритмов по качеству необходимо проверить гипотезу $\mathcal{H}_0 : \Delta\mu = 0$ против альтернативы $\mathcal{H}_1 : |\Delta\mu| > 0$ по выборкам

$(x_1^1, \ldots, x_n^1)$ и $(x_1^2, \ldots, x_n^2)$, показывающим значения их метрик, полученных в эксперименте.

Проведите статистическое моделирование для сравнения эффективности нескольких распространенных статистических критериев в задаче различения алгоритмов по качеству.

- В качестве множества критериев рассмотрите: критерий Вальда, критерий Стьюдента, критерий Манна-Уитни-Уилкоксона, критерий знаков и критерий перестановок. Воспользуйтесь известными в библиотеках реализациями (например, критерий Манна-Уитни-Уилкоксона реализуется функцией `scipy.stats.mannw x2, alternative='two-sided')`.
- В качестве множества постановок рассмотрите ситуации, когда X_1 и X_2 имеют:
  - одинаковый тип распределения и равные стандартные отклонения;
  - одинаковый тип распределения, но неравные стандартные отклонения;
  - различные типы распределения и равные стандартные отклонения;
  - различные типы распределения и неравные стандартные отклонения.
- В качестве типов распределения рассмотрите следующие: стандартное нормальное распределение, логнормальное распределение, распределение Коши (с «тяжелыми хвостами») на отрезке $[-3, 3]$. Все распределения отмасштабируйте так, чтобы их среднее $\mu = 0$, стандартное отклонение $\sigma = 1$. При рассмотрении различных стандартных отклонений положите $\sigma_2 = 2\sigma_1$.

- Проведите следующие эксперименты. Все эксперименты необходимо провести моделированием Монте-Карло с числом повторений $N_r = 10^3$, для каждого критерия, каждой постановки и размеров выборок $N_s \in \{1, 2, 3, 4, 5, 10, 20, 30, 40, 50, 100\}$.
  - Измерение вероятности ложной тревоги: зафиксируйте $\alpha = 0.05$ и при верной $\mathcal{H}_0$ подсчитайте долю случаев, в которых была отклонена гипотеза $\mathcal{H}_0$.
  - Измерение мощности теста: зафиксируйте $\alpha = 0.05$ и при верной $\mathcal{H}_1$ подсчитайте долю случаев, в которых была отклонена гипотеза $\mathcal{H}_0$. При этом размер сдвига $\Delta\mu$ варьируйте в диапазоне от 0 до 3 с шагом 0.1.

Требования к оформлению результатов в этой задаче:

- Должны быть представлены графики зависимостей вероятности ошибки I рода от размера выборки N_s для каждой постановки (при этом на одном и том же графике должны быть представлены кривые для каждого критерия). Сгруппируйте графики по типу рассматриваемой постановки (например, в разделе «одинаковый тип распределения и равные стандартные отклонения» должно быть 3 графика, на каждом по 5 кривых, и т.д.).
- Должны быть представлены графики зависимостей мощности критерия от размера выборки N_s для каждой постановки (при этом на одном и том же графике должны быть представлены кривые для каждого критерия). Сгруппируйте графики по типу рассматриваемой постановки.
- К отчету должен быть приложен исходный код, реализующий сравнение.

3. (3 балла) Во время Второй Мировой войны в лондонской газете выложили карту падения бомб V-1 и V-2 в Центральном Лондоне. Горожане обратили внимание, что есть районы с более высокой кучностью падения бомб, а есть кварталы, которые вообще не были затронуты. Из этих наблюдений жители сделали два вывода: (i) немцы обладают высокоточными бомбами; (ii) в незатронутых кварталах скорее всего живут немецкие шпионы.

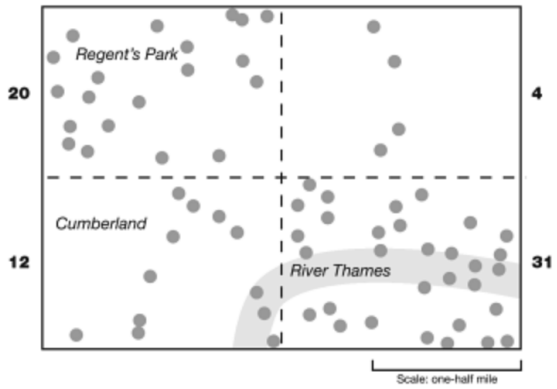

Figure 1: Adapted from Gilovich (1991)

Вам дан датасет с координатами падения бомб во Второй Мировой войне и ваша задача сказать насколько правы были жители.

- Скачайте данные: https://github.com/SchattenGenie/hse-stats-course-2019/blob/master/homeworks/hw_2/v2_bombing_london.csv;
- В файле представлены данные с падения бомб для всего Лондона, включая предместья, поэтому необходимо сделать очистку данных, выбрав только Центральный Лондон(или по картинке выше подбирайте на глаз или попробуйте отфильтровать координаты с помощью np.percentile);
- Придумайте тест для различия следущих гипотез: $\mathcal{H}_0$ – бомбы падали равномерно *на плоскости* v.s. $\mathcal{H}_1$ – бомбы падали неравномерно;

- (1 балл) Реализуйте и примените тест для $\alpha = 0.05$;
- (1 балл) Исследуйте (найдите) эмпирическую ошибку I рода;
- (1 балл) Рассмотрим следующее распределение на двумерной плоскости: $[\text{Beta}(x; 1+\epsilon, 1+\epsilon) \times \text{Beta}(y; 1+\epsilon, 1+\epsilon)]$.

  Проварьируйте $\epsilon \in [-1, 2]$ и постройте эмпирическую ошибку II рода для вашего теста;
- Сделайте выводы.

4. (2 балла) Вы попадаете на остров, на котором живёт племя туземцев. Вы считаете, что племя – часть определённого народа с характерной долей 1-й группы крови. Какое минимальное число n анализов крови нужно сделать, чтобы подтвердить вашу гипотезу с погрешностью не более $\Delta = 0.02$, с вероятностью $q \geqslant 0.95$? Рассмотрите отдельно общий случай (людей на острове бесконечно много).

Бонусные задачи