ПСМО ФКН ВШЭ, 3 курс, 1 модуль

Задание 2. Параметрическое оценивание и дистанции.

Прикладная статистика в машинном обучении, осень 2018

Время выдачи задания: 10 октября (среда).

Срок сдачи: **24 октября (среда), 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. (1 балл) Наблюдаемый самолёт характеризуется расстоянием до наблюдателя, r, и углом наблюдения, θ. Пусть есть m измерений R и Θ, найдите вариацию высоты самолёта, вычисляемую по формуле $Y = R \sin \Theta$. Если R фиксирована, когда достигается максимальная вариация Y?

2. (2 балла) Напишите программу, оценивающую дистанции Кульбака-Лейблера, χ^2 и полной вариации между $\mathcal{N}(0,1)$ и $\frac{1}{2}\left(\mathcal{N}(\mu,1) + \mathcal{N}(-\mu,1)\right)$. Постройте график зависимости дистанций от значений μ на промежутке $[-1;1]$.

3. (3 балла) Для случайной величины $X \sim \mathcal{N}(\mu, \sigma^2)$ и функции $g(x) = \exp(x)/(1+\exp(x))$

   - (1 балл) Получите выражение для среднего и дисперсии $g(x)$ с помощью дельта-метода.
   - (1 балл) Получите такие же оценки с помощью бутстрапа.
   - (1 балл) постройте график зависимости дисперсии от значений параметров. Сделайте вывод.

4. (4 балла) У Вас есть две монеты, P и Q одна из которых имеет дефект. Из-за этого дефекта вероятность выпадения орла в монете P выше на 2ϵ, чем вероятность выпадения решки. Для второй монеты Q эти вероятности равны 0.5. Вы можете подбрасывать монету и смотреть на результат. Существует алгоритм $A(x_1, .., x_m) \to 0; 1$, который говорит, является ли монета деффектной ($A = 0$) или настоящей ($A = 1$) на основании m независимых подбрасываний. С помощью неравенства Пинскера и свойств дивергенции Кульбака-Лейблера найдите минимальное m, для которого A сможет полу-

чить ответ с вероятностью больше 90% $P_{x \in Q}(A(x) = 1) > 0.9$, $P_{x \in P}(A(x) = 1) > 0.9$.

(a) (1 балл) Докажите, что для любых распределений $\tilde{P}$ и $\tilde{Q}$ над U и функции $f(x) : U \to [0; B]$ выполнено:

$$|\mathbb{E}_{\tilde{P}}[f(x)] - \mathbb{E}_{\tilde{Q}}[f(x)]| \leq \frac{B}{2}||\tilde{P} - \tilde{Q}||,$$

где $||.||$ – расстояние полной вариации. Для доказательства можно использовать дискретные распределения.

(b) (1 балл) Воспользовавшись результатом предыдущего пункта, найдите нижнее ограничение на расстояние полной вариации между P^m и Q^m, считая $f = A$.

(c) (1 балл) Найдите верхнюю границу $KL(P||Q)$ с точностью до ϵ^2, считая $\epsilon < 0.25$.

(d) (1 балл) Используя свойства КЛ дивергенции большого количества семплов и неравенство Пинскера, найдите нижнюю оценку на m через ϵ с точностью до $1/\epsilon^2$

Бонусные задачи

1. (1 балл) Найдите выражение для дивергенции Кульбака-Лейблера между двумя нормальными распределениями $\mathcal{N}(\mu_1; \sigma_1)$ и $\mathcal{N}(\mu_2; \sigma_2)$.

2. (1 балл) (По стопам семинара про натуральный градиент) В семинаре матрица Фишера считалась эмпирически по выборке. Этот подход имеет преимущество в случаях, когда плотность распределения неизвестна или сложно вычислима, но посчитать градиент достаточно просто (к примеру, когда распределение порождается нейронной сетью). Альтернативно, если плотность распределения известна, то матрица Фишера вычислима аналитически. В этой задаче мы сравним эти два подхода.

   Вам предлагается проделать следующие шаги:

   (a) Сгенерировать две выборки размером $N = 50$ и $N = 2000$ из двумерного нормального распределения с параметрами: $\mu_0 = (2, -6) = (\mu_1, \mu_2)$ и $\Sigma_0 = \begin{pmatrix} 5 & 0.8 \\ 0.8 & 2 \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \rho \\ \rho & \sigma_{22} \end{pmatrix}$.

   (b) Вывести аналитическое выражение матрицы Фишера для вектора параметров: $\theta = (\mu_1, \mu_2, \sigma_{11}, \rho, \sigma_{22})$;

   (c) написать код для натурального градиентного спуска с эмпирической матрицей Фишера и с аналитической матрицей Фишера для задачи максимизации функции правдоподобия;

   (d) сравнить для двух выборок скорость сходимости по метрике KL для двух альтернативных алгоритмов. В качестве начальной точки взять: $\mu = (0, 0)$ и $\Sigma = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

   (e) сделать выводы по полученным результатам.