

## Задачи первичного отборочного теста по направлению DataScience

(матанализ, линейная алгебра, аналитическая геометрия, теория вероятностей, классическая статистика, машинное обучение, машинное зрение)

2022-05-18

1. Дефектоскопист фиксирует поверхностный дефект на прокате с вероятностью 0,8 и принимает некий артефакт за дефект с вероятностью 0,1. Опыт показывает, что в 15 % случаев на поверхности металла находятся несущественные артефакты (пятна, грязь и т.п.). Дефектоскопист, изучая подозрительное пятно, принял решение о наличии в этом месте дефекта. Определить вероятность того, что это пятно действительно является дефектом.
2. Обычная колода карт насчитывает 52 листа, среди которых имеются 4 туза. Колода тщательно тасуется, после чего открываются верхние карты одна за другой до появления первого туза. Подсчитывается номер карты, которая оказалась первым тузом. Этот процесс повторяется несколько раз.

Численно смоделируйте этот эксперимент. Постройте гистограмму распределения количества карт до выпадения первого туза. Среди какого количества карт, лежащих сверху колоды, заведомо находится первый туз в половине случаев? Как называется эта величина? Модель можно представить в виде кода на Python.

3. Случайная величина  $X$  имеет нормальное распределение с параметрами  $\mu=10$ ,  $\sigma^2=2$ . В ходе экспериментов получили выборку из 10 значений этой случайной величины. Какое распределение будет иметь среднее значение величины  $X$ ? Выпишите параметры этого распределения. С какой вероятностью среднее выборочное окажется больше 12.5?

4. Решить уравнение 
$$\begin{vmatrix} 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1-x & 0 & 0 & \dots & 0 \\ 0 & 0 & 2-x & 0 & \dots & 0 \\ 0 & 0 & 0 & \ddots & \dots & 0 \\ \dots & \dots & \dots & \dots & \ddots & \dots \\ 0 & 0 & 0 & 0 & 0 & n-x \end{vmatrix} = 0.$$
 Символ  $||$  означает определитель.

5. Решить матричное уравнение  $\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} X = \begin{pmatrix} -1 & 0 \\ 2 & 1 \end{pmatrix}$ .  $X$  — матрица. Письменный ход решения надо представить в виде скана (фото).
6. На плоскости  $(X, Y)$  дана окружность радиуса  $R$  с центром в точке  $(X_0, Y_0)$ . Дана прямая, лежащая в этой же плоскости, наклоненная к оси  $X$  под углом  $45^\circ$ . Известно, что прямая касается окружности. Напишите координаты точек касания.
7. Функция стоимости от единственного технологического параметра  $x \geq 0$  описывается выражением  $cost(x) = (x^2 - x)e^x$ . При каком  $x$  стоимость примет минимальное значение и каково значение минимума?
8. Щупом измерили координаты четырех точек поверхности некоторой плиты на инспекционном столе и получили следующие значения координат:  $A_1 = (1; 0; 2)$ ,  $A_2 = (2; -1; 0)$ ,  $A_3 = (0; 0; 0)$ ,  $A_4 = (1; -1; 0)$ . Является ли поверхность плиты плоской (измерения будем считать точными)? <sup>Ω</sup>Какую из четырех точек следует отбросить в первую очередь, если поверхность окажется все же неплоской?
9. На производстве  $N$  задумали изготавливать продукцию с выходом годного не менее 95 %. Известно, что уровень качества там определяется 5 независимыми технологическими факторами, имеющими одинаковый шанс “испортить” металл в случае возникновения их от-

---

<sup>Ω</sup>Вопросы, отмеченные знаком  $\Omega$ , являются дополнительными.

клонений в ходе технологического процесса. Определить, при какой (максимальной) вероятности возникновения отклонений отдельного фактора удастся выполнить поставленную цель.

10. В следующей таблице указана относительная оценка дефектности и степени легирования 8 марок сталей, занумерованных числами от 1 до 8, причем менее дефектным маркам соответствует меньший ранг, менее легированным — меньшее значение ранга.

№ марки стали	1	2	3	4	5	6	7	8
Дефектность	5	2	1	7	8	4	6	3
Легированность	4	3	2	1	7	6	5	8

Некто считает, что дефектность и легированность сталей независимы в противоположность альтернативному мнению о том, что они положительно коррелированы. Проверьте эту гипотезу на уровне значимости  $\alpha = 0,1$ .

11. <sup>Ω</sup>Бинарная классификация. Исходные данные в прилагаемом файле Users\_behavior.csv. Каждый объект в наборе данных — это информация о поведении одного пользователя за месяц. Известно:

- calls — количество звонков,
- minutes — суммарная длительность звонков в минутах,
- messages — количество sms-сообщений,
- mb\_used — израсходованный интернет-трафик в Мб,
- is\_ultra — каким тарифом пользовался в течение месяца («Ультра» — 1, «Смарт» — 0).

Необходимо построить модель классификации для определения того, является ли тариф «Ультра» или «Смарт».

12. <sup>Ω</sup>Машинное зрение. Файл с изображением текста прилагается — Hymn\_text.jpg. Подсчитать количество вхождений какого-либо одного символа в данном тексте. <sup>ΩΩ</sup>Подсчитать частоту вхождений всех символов. Представить код функции, возвращающей словарь, в котором ключ — символ, значение — число вхождений символа. На собеседовании дать пояснение по использованному методу с указанием ссылок на литературу.

### Оценка задач

Номер задачи	Баллы
1	3
2	2
3	2
4	1
5	2
6	1
7	1
8	2 (+3 <sup>Ω</sup> )
9	2
10	2
11	4 <sup>Ω</sup>
12	4 <sup>Ω</sup> (+3 <sup>ΩΩ</sup> )