Задача 4.2

Агафонов Артём

- Будем проверять гипотезу равенства средних для классификаторов. Если получится так, что мы не отклоним эту гипотезу, то исследователь не сможет утвержать, что одна из моделей лучше других.

$$4 \text{ связанные выборки } X_i = \{X_i^j\}_{j=1}^6, \ i \in [1, 4]$$
$$H_0: \ X_i \text{ пришли из распределений с одинаковым средним}$$
$$H_1: \text{двусторонняя альтернатива}$$

- Для решения задачи мы будем попарно считать раность $R_i - R_j$, где R_i – сумма рангов элементов выбоки i-ого классификатора. Для того, чтобы их вычислить нам необходимо объеденить все выборки и составить вариационный ряд. В таблице ниже представлены исходные данные. В скобках указаны ранги элементов.

  Критерий, которым мы будем пользоваться, называется rank sum difference test (тест 55 [Kanji, 2006])

a_1	a_2	a_3	a_4
86 (21)	50 (12)	93 (23)	13 (3)
85 (20)	74 (19)	55 (15)	35 (7.5)
53 (14)	92 (22)	58 (17)	51 (13)
44 (11)	41 (10)	56 (16)	37 (9)
2 (1)	18 (5)	99 (24)	26 (6)
5 (2)	68 (18)	35 (7.5)	17 (4)
$R_1 = 69$	$R_2 = 86$	$R_3 = 102.5$	$R_4 = 42.5$

$$|R_1 - R_2| = 17 \ |R_1 - R_3| = 33.5 \ |R_1 - R_4| = 26.5$$
$$|R_2 - R_3| = 16.5 \ |R_2 - R_4| = 43.5 \ |R_3 - R_4| = 60$$

- Нулевое распределение табличное. Критические значения этого критерия можно найти в таблице 23 [Kanji, 2006].

- Решающее правило выглядит следующим образом. Если одна из разностей превосходит критическое значение, то гипотеза H_0 отвергается.

- Для значений $n = 6$ и $K = 4$ и уровня значимости $\alpha = 0.05$ критическое значение равно 62.9. Для наших данных получилось, что

$$\max_{i,j \in [1,4]} |R_i - R_j| < 62.9.$$

  Это означает, что мы не отвергаем гипотезу H_0. Исследователь не может утвержать, что одна из моделей лучше другой.

Замечание. Мы решали задачу в предположении, что выборки связанные, ведь нам ничего о них не известно. В противном случае можно было бы предложить другое решение задачи. Например, если качество – accuracy и мы знаем размеры выборок и то, что они iid (такие предположения вполне логичны, ведь обычно классификаторы работают для конкретных задач, где считается, что объекты идут независимо из одного распредения), то можно было бы воспользоваться критерием однордности χ^2 (условие $n > 40$ скорее всего было бы выполнено, иначе выборки были бы слишком маленькими для обучения).

References

[Kanji, 2006] *Kanji J. K.* 100 statistical tests — 2006.