

Задача 4.2

Кубраков Никита

March 2020

1 Задача 4.2

Рассмотрим некоторую задачу классификации. Пусть задано качество 4 моделей a_1, a_2, a_3, a_4 . Качество полученных моделей показано в таблице.

Исследователю требуется выбрать наилучшую модель. Для выбора лучшей модели исследовать требуется попарно сравнить среднее значение качества всех моделей. Может ли исследователь утверждать что какая-то из моделей лучше другой?

Требуется:

1. записать задачу формально;
2. предложить статистику для решения данной задачи;
3. записать нулевое распределение данной статистики;
4. записать явно правило принятия решения на основе статистики и нулевого распределения для обеспечения уровня значимости $\alpha = 0.05$; проверить гипотезу по записанному критерию, для данных из условия. Противоречат ли они гипотезе?

2 Решение

1. Так как мы не знаем, из какого распределения вектора результатов, для каждой модели, а попарно вектора двух моделей связаны, то удобно воспользоваться критерием знаковых рангов Уилкоксона для связанных выборок.

Формально:

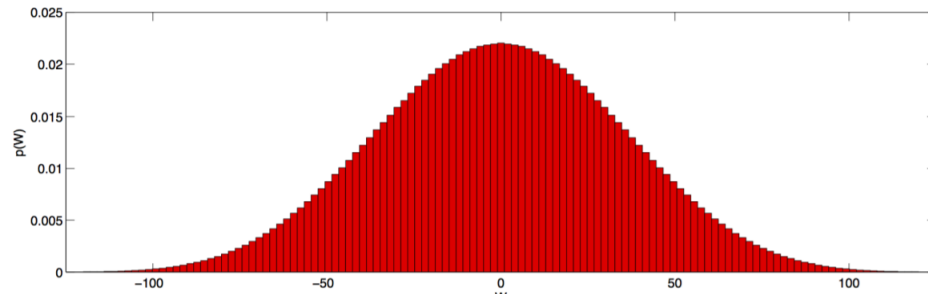
выборки: одного размера, связанные, элементы с одинаковым индексом не совпадают

нулевая гипотеза: $med(X_1 - X_2) = 0$

альтернатива: $med(X_1 - X_2) \neq 0$

статистика: $\sum_{i=1}^n rank(|X_{1i} - X_{2i}|) \cdot sign(X_{1i} - X_{2i})$

нулевое распределение: табличное



2. Теперь пользуясь всем вышесказанным нахожу попарные p-value для нулевых гипотез об одинаковом качестве моделей:

```
a1 a2 0.463071015014588
a1 a3 0.2071604489114608
a1 a4 0.463071015014588
a2 a3 0.7531523647659145
a2 a4 0.07473549830588248
a3 a4 0.027707849358079864
```

Как видно везде кроме последней пары на нашем уровне значимости данные не противоречат гипотезе. Но мы находимся в задаче о множественной проверке. Поэтому воспользуемся различными поправками, для установления истинного p-value:

```
bonferroni
[False False False False False]
[1.         1.         1.         1.         0.44841299 0.1662471 ]
-----
holm
[False False False False False]
[1.         0.8286418 1.         1.         0.37367749 0.1662471 ]
-----
fdr_bh
[False False False False False]
[0.55568522 0.4143209 0.55568522 0.75315236 0.22420649 0.1662471 ]
-----
fdr_by
[False False False False False]
[1.         1.         1.         1.         0.54930591 0.40730539]
-----
holm-sidak
[False False False False False]
[0.84520727 0.60486889 0.84520727 0.84520727 0.32184417 0.15514792]
-----
```

Какой бы поправкой в данном случае мы не воспользовались, итог один - мы не можем говорить о том, что какой-то классификатор лучше какого-то другого. Значит либо исследователю придется собрать больше данных, либо не выбирать лучшую модель, а построить ансамбль:)