

Классификация вин

Версия 2

Задача. Измеряются 13 характеристик химического состава вина. Необходимо по значениям имеющихся переменных определить тип вина.

Имеются результаты химического анализа трех сортов вина, grown in the same region in Italy but derived from three different cultivars.

Сорт вина указан в трех столбцах класс_1, класс_2 и класс_3. Если в первом столбце (класс_3) стоит единица, то наблюдение соответствует виду вина третьего типа, если во втором столбце (класс_2) стоит единица, то наблюдение соответствует виду вина второго типа, если в третьем столбце (класс_1) стоит единица, то наблюдение соответствует виду вина первого типа. Таким образом, столбцы дублируют друг друга.

Задача приобрела большую известность и активно обсуждалась. Как оказалось, статистические методы могут различать вина лучше опытного профессионального знатока вин (сомелье). Эффект после начальной публикации оказался хоть и ниже, чем от проигрыша Каспаровым (на тот момент чемпионом мира) матча шахматной программе Дип Блю» (Deep Blue) но тем не менее сопоставимым.

Список переменных:

- 1) Alcohol (содержание алкоголя)
- 2) Malic acid (яблочная кислота)
- 3) Ash (зола)
- 4) Alcalinity of ash
- 5) Magnesium (магний)
- 6) Total phenols (общее содержание фенола (карболовой кислоты))
- 7) Flavanoids (ароматические вещества)
- 8) Nonflavanoid phenols
- 9) Proanthocyanins
- 10) Color intensity (интенсивность цвета)
- 11) Hue (окраска? красители?)
- 12) OD280/OD315 of diluted wines
- 13) Proline (пролин)

Число наблюдений – 178

Число переменных – 13, все измерены в количественной (непрерывной) шкале.

В задаче рекомендуется (но не обязательно) провести сокращение числа переменных.

Источник задачи: UCI Machine learning Database
<ftp://ftp.ics.uci.edu/pub/machine-learning-databases/wine>

Обсуждение задачи:

<ftp://ftp.ics.uci.edu/pub/machine-learning-databases/wine>

Данные обсуждаются в книге

Principe Jose C., Euliano Neil R., Lefebvre W. Curt

Neural and Adaptive Systems: Fundamentals through Simulations