

Практическое задание №3 по курсу МИАД.
Асириян Александр, 524 группа.

1. 4.) ...Откройте список переменных на этом узле и поставьте роль для переменной с префиксом PCTL_ Stratification. Зачем это делается?
Ответ: для сохранения исходного распределения.
2. 5.) ...Какие значения ASE получились на тренировочном, валидационном и тестовом наборе данных?
Ответ: тренировочный – 0.50, валидационный – 0.52, тестовый – 0.56.
3. 6.) После узла Transform variables и перед узлом MBR вставьте узел Variable clustering (Вариант II). Сколько переменных теперь передается на вход MBR? Как изменились значения ASE в результате? Почему это могло произойти?
Ответ: 6, 4 из которых имеют роль Input. ASE на тренировочном наборе выросла на 0.06 (0.56), на валидационном на 0.05 (0.57) и на тестовом на 0.07 (0.63). Слишком сильно снизилась размерность, из-за чего MBR стал работать хуже.
4. 7.) Какая модель показывает лучшее качество на тестовом наборе? Какие переменные не вошли в результирующую регрессионную модель? На каком шаге была выбрана лучшая модель (см. график Iteration plot (View->Model)? Посмотрите в журнале (раздел Output) какая из вошедших переменных наименее важная с точки зрения t-статистики?
Ответ: модель регрессии лучше. Не вошли chlorides, citric_acid, total_sulfur_dioxide. На 8-м шаге. Наименее важная – fixed_acidity.
5. 8.) Добавьте узел Transform variables перед узлом Regression и сделайте настройку Interval Inputs – Maximum Normal. Как изменились результаты регрессионной модели с точки зрения числа степеней свободы и значения ASE на тестовом наборе?
Ответ: было 9 степеней свободы и ASE = 0.56. Степени свободы не изменились, ASE уменьшилась на 0.01 (0.55).
6. 9.) Отредактируйте уравнение регрессии, чтобы оно стало полным полиномом второй степени, в разделе Equation задайте Polynomial Terms = Yes, Two-factor interactions = yes. Normal. Как изменились результаты регрессионной модели с точки зрения числа степеней свободы и значения ASE на тестовом наборе?
Ответ: число степеней свободы увеличилось на 8 (17), а ASE уменьшилась на 0.02 (0.53).
7. 10.) ... Сколько компонент было выбрано? Какие переменные были отобраны по абсолютным значениям коэффициентов (II Вариант). Какое значение ASE модель показывает на тестовом наборе?
Ответ: было выбрано 7 компонент. Были отобраны 5 переменных: alcohol, density, pH, sugar, volatile_acidity. ASE равняется 0.56.

8. 11.) ...Какое значение ASE модель показывает на тестовом наборе? Какая переменная была отобрана на первом шаге? Сколько всего переменных было отобрано? Обратите внимание на компоненту Iteration Plot в результатах. Если бы вместо критерия отбора модели вашего варианта использовался критерий ASE на валидационном наборе, то было бы отобрано больше или меньше переменных?

Ответ: ASE = 0.56. Первая переменная – alcohol. Были отобраны все (11). Было бы выбрано столько же переменных.