

# Лабораторная работа №4: Методы снижения размерности. Регуляризация логистической регрессии.

## Упражнение 3

1. Данные своего варианта (см. таблицу ниже) разделить на выборку для построения моделей (85%) и отложенные наблюдения (15%). Оставить в таблице только указанные в варианте переменные. Отложенные наблюдения использовать только для прогноза по лучшей модели.
2. Провести предварительный и корреляционный анализ данных с помощью статистик и графиков из этой лабораторной.
3. Снизить размерность пространства объясняющих переменных методом, указанным в варианте. Обосновать количество главных компонент. Построить модель логистической регрессии на преобразованном пространстве объясняющих переменных. Оценить точность модели методом перекрёстной проверки.
4. Провести регуляризацию модели логистической регрессии методом, указанным в варианте. Подобрать оптимальное значение гиперпараметра методом перекрёстной проверки. Построить график сжатия параметров модели.
5. Выбрать наиболее точную модель из полученных в пунктах 3 и 4, сделать по ней прогноз на отложенные наблюдения, оценить точность этого прогноза.

## Варианты

Номер варианта – номер студента в списке. Студент под номером 21 берёт вариант 1, под номером 22 – 2, и т.д.

В качестве ядра генератора случайных чисел (в частности, для разделения данных на выборку для построения моделей и отложенные наблюдения) используйте номер своего варианта. Наборы данных и справочники к ним выложены здесь: <https://github.com/ania607/ML/tree/main/data>.

Номер варианта	Данные	Зависимая переменная	Объясняющие переменные	Метод снижения размерности	Метод регуляризации
1	winequality-red_for_lab	Y	citric.acid, residual.sugar, chlorides, free.sulfur.dioxide, total.sulfur.dioxide,	PCR	ридж-регрессия

			density, pH, sulphates, alcohol		
2	winequality-red_for_lab	Y	citric.acid, residual.sugar, chlorides, free.sulfur.dioxide, total.sulfur.dioxide, density, pH, sulphates, alcohol	PLS	лассо-регрессия
3	winequality-red_for_lab	Y	fixed.acidity, volatile.acidity, residual.sugar, chlorides, free.sulfur.dioxide, total.sulfur.dioxide, pH, sulphates, alcohol	PCR	ридж-регрессия
4	winequality-red_for_lab	Y	fixed.acidity, volatile.acidity, residual.sugar, chlorides, free.sulfur.dioxide, total.sulfur.dioxide, pH, sulphates, alcohol	PLS	лассо-регрессия
5	winequality-white_for_lab	Y	fixed.acidity, volatile.acidity, citric.acid, residual.sugar, chlorides, free.sulfur.dioxide, pH, sulphates, alcohol	PCR	ридж-регрессия
6	winequality-white_for_lab	Y	fixed.acidity, volatile.acidity, citric.acid, residual.sugar, chlorides, free.sulfur.dioxide, pH, sulphates, alcohol	PLS	лассо-регрессия
7	winequality-white_for_lab	Y	fixed.acidity, volatile.acidity, citric.acid, chlorides, total.sulfur.dioxide, density, pH, sulphates, alcohol	PCR	ридж-регрессия
8	winequality-white_for_lab	Y	fixed.acidity, volatile.acidity, citric.acid, chlorides, total.sulfur.dioxide, density, pH, sulphates, alcohol	PLS	лассо-регрессия
9	winequality-red_for_lab	Y	все, кроме quality и Y	PCR	ридж-регрессия
10	winequality-red_for_lab	Y	все, кроме quality и Y	PLS	лассо-регрессия
11	winequality-white_for_lab	Y	все, кроме quality и Y	PCR	ридж-регрессия

12	winequality- white_for_lab	Y	все, кроме quality и Y	PLS	лассо- регрессия
13	default_of_credit- card_clients	Y	все, кроме Y	PCR	ридж- регрессия
14	default_of_credit- card_clients	Y	все, кроме Y	PLS	лассо- регрессия
15	default_of_credit- card_clients	Y	все, кроме: Y, PAY_4, PAY_5, PAY_6, BILL_AMT4, BILL_AMT5, BILL_AMT6, PAY_AMT4, PAY_AMT5, PAY_AMT6	PCR	ридж- регрессия
16	default_of_credit- card_clients	Y	все, кроме: Y, PAY_4, PAY_5, PAY_6, BILL_AMT4, BILL_AMT5, BILL_AMT6, PAY_AMT4, PAY_AMT5, PAY_AMT6	PLS	лассо- регрессия
17	default_of_credit- card_clients	Y	все, кроме: Y, PAY_4, PAY_5, PAY_6, BILL_AMT4, BILL_AMT5, BILL_AMT6, PAY_AMT1, PAY_AMT2, PAY_AMT3, PAY_AMT4, PAY_AMT5, PAY_AMT6	PCR	ридж- регрессия
18	default_of_credit- card_clients	Y	все, кроме: Y, PAY_4, PAY_5, PAY_6, BILL_AMT4, BILL_AMT5, BILL_AMT6, PAY_AMT1, PAY_AMT2, PAY_AMT3, PAY_AMT4, PAY_AMT5, PAY_AMT6	PLS	лассо- регрессия
19	default_of_credit- card_clients	Y	все, кроме: Y, PAY_3, PAY_4, PAY_5, PAY_6, BILL_AMT3, BILL_AMT4, BILL_AMT5, BILL_AMT6	PCR	ридж- регрессия
20	default_of_credit- card_clients	Y	все, кроме: Y, PAY_3, PAY_4, PAY_5, PAY_6, BILL_AMT3, BILL_AMT4, BILL_AMT5, BILL_AMT6	PLS	лассо- регрессия