

# Проверочная работа 2

Факультатив «Введение в анализ данных и машинное обучение на Python»

21 декабря 2019

## Задание 1

Пусть представлены следующие данные:

$$Y = \begin{pmatrix} 6 \\ 12 \\ 24 \end{pmatrix}, X = \begin{pmatrix} 1 & 6 \\ 1 & 3 \\ 1 & 6 \end{pmatrix}$$

- а) Оценивается следующая модель:

$$Y = X\beta + u,$$

где  $u$  – случайная ошибка.

Найдите оценки коэффициентов модели. Найдите прогноз модели.

- б) Нанесите выборку на диаграмму рассеяния. Постройте оценённую линию регрессии.
- в) Определите направление корреляции между  $X$  и  $Y$ .
- г) Найдите выборочное среднее и выборочную дисперсию неконстантного признака.
- е) Найдите коэффициент детерминации данной регрессии.

## Задание 2

- а) Запишите уравнение парной регрессии. Запишите уравнение оценённой линии в этой регрессии.
- б) Запишите уравнение множественной регрессии в общем виде. Под каждым элементом (матрицей или вектором) подпишите его размер.
- в) Является ли линейной регрессией следующая модель:  $Y = \beta_0 + \beta_1 X_1 X_2 X_3 + u$ ?
- г) Является ли линейной регрессией следующая модель:  $Y = \beta_0 + \beta_1^2 X_1 + \beta_2^3 X_2 + u$ ?
- е) Является ли линейной регрессией следующая модель:  $Y = \beta_0 + \beta_1 X_1^5 + u$ ?

## Задание 3

- а) Что такое мультиколлинеарность признаков? К чему она может привести? Что можно сделать для устранения этой проблемы?
- б) Опишите (по смыслу, без формул), как регуляризация помогает справляться с переобучением линейной регрессии.

## Задание 4

- a) Опишите своими словами, что такое машинное обучение. В чём основная цель машинного обучения?
- b) Какие три типа задач машинного обучения в зависимости от наличия целевой переменной выделяют?
- c) Приведите пример задачи многоклассовой классификации.
- d) Как происходит обучение алгоритма (модели) машинного обучения?
- e) Чем функционал качества отличается от метрики качества?
- f) Что такое недообучение? На одном графике схематически изобразите недообученную и корректно обученную модель.