

## Регрессионный анализ.

- 1 (3 балла) Рассмотрим линейную модель  $Y = \theta + \varepsilon$ , где  $E\varepsilon = 0$ . Докажите, что значение  $\theta$ , при котором величина  $\|Y - \theta\|^2$  минимизируется при ограничении  $A\theta = 0$ , где  $A$  – известная матрица размера  $q \times n$  ранга  $q$ ,  $q < n$ , равно

$$\hat{\theta} = (I_n - A^T(AA^T)^{-1}A)Y.$$

- 2 (4 балла) Выданы данные о связи между уровнем антигена (lpsa), специфичного при наличии рака простаты, с рядом клинических показателей (столбцы 1-8), которые были измерены у мужчин непосредственно перед проведением операции. Выбрав наилучшую модель линейной регрессии, предсказать lpsa на тестовой выборке. Описать и объяснить проделанные процедуры.
- 3 (5 баллов) Мистер Фаттахов – большой любитель дышать свежим воздухом и совершать пешие прогулки. Часто компанию мистеру Фаттахову составляют его друзья Фадеев, Калдеев и Пепермалдеев, и тогда приятность, а также длительность прогулки существенно возрастают. Найти зависимость между длиной прогулки мистера Фаттахова и его компанией, набором предметов для прогулки и длительностью глубокой фазы его послеобеденного сна. Описать и объяснить проделанные процедуры.
- 4 (6 баллов) Выданы данные  $\{(y_i, x_{ij}), i = 1, \dots, n+q, j = 1, \dots, k\}$ , причем  $y_{n+1}, \dots, y_{n+q}$  неизвестны. Используя пройденные методы регрессионного анализа, в рамках линейной регрессионной модели произвести отбор признаков по первым  $n$  объектам и предсказать значения откликов объектов с номерами  $n + 1, \dots, n + q$ . Описать и объяснить проделанные процедуры.
- 5 (6 баллов) Выданы данные  $\{(y_i, x_{ij}), i = 1, \dots, n+q, j = 1, \dots, k\}$ , причем  $y_{n+1}, \dots, y_{n+q}$  неизвестны. Используя стандартную и робастные регрессионные модели, произвести отбор признаков по первым  $n$  объектам и предсказать значения  $y_{n+1}, \dots, y_{n+q}$ . Какая модель дала наилучший результат и почему? Можно ли пользоваться стандартными процедурами, такими, как критерий Фишера и информационные критерии, для отбора признаков в робастных моделях?