

## Практика 7. Анализ регрессионных остатков и погрешности регрессионных моделей характеристик насосных агрегатов

### Часть 1. Анализ регрессионные остатков

1. Сформулировать задачу анализа регрессионных остатков критерием серий (Вальда-Вилкоксона).
2. Для моделей всех порядков из предыдущего ДЗ по напорной характеристике и характеристике мощности:
  - a. Построить графики регрессионных остатков по номерам экспериментов (не по расходу!)
  - b. Применить функцию `runstest` для анализа остатков моделей всех порядков из предыдущего ДЗ.
3. Выбрать модель напорной характеристики минимального порядка, для которой гипотеза о наличии закономерности в регрессионных остатках отклоняется.
4. Выбрать модель характеристики мощности минимального порядка, для которой гипотеза о наличии закономерности в регрессионных остатках отклоняется.

### Часть 2. Скользящий контроль

1. Построить оценки погрешности модели на скользящем контроле LOO (leave-one-out):
  - a. Оценка дисперсия шума  $\hat{D}_{LOO}$ , построенная по экзаменационным выборкам (как в лекциях).
  - b. Вывести формулу  $RSS_{LOO}$  для построения по экзаменационным выборкам (по аналогии с оценкой дисперсии  $\hat{D}_{LOO}$ ). То же самое для  $TSS_{LOO}$ , учтя что  $RSS_0 = TSS$ . Затем по полученным значениям вычислить  $R^2_{LOO}$  по экзаменационным выборкам.
  - c. Коэффициент детерминации  $R^2$ , усредненный по обучающим выборкам
  - d. Несмещенная оценка дисперсии шума, построенная по обучающим выборкам (для сравнения с  $\hat{D}_{LOO}$ ).
2. Для каждого разбиения выборки на скользящем контроле построить 95%-доверительный интервал предсказания нового значения. Вывести количество и долю попаданий нового значения в 95%-доверительный интервал прогноза.

### Вопросы к защите (20.05.2017)

1. Статистический критерий серий (Вальда-Вольфовица).
  - a. Случайный ряд, ряд с трендом, ряд с периодичностью. Исходя из определения серии, указать, для какого типа ряда их количество  $N_S$  больше или меньше.
  - b. Формулировки основной и альтернативной гипотезы.
  - c. Эскизы распределений статистики  $N_S$  в условиях основной и альтернативной гипотез.
  - d. Формула стандартизированной статистики  $z$
  - e. Распределения статистики  $N_S$  и стандартизированной статистики  $z$  в условиях основной гипотезы (точная форма распределения).
2. Анализ регрессионных остатков
  - a. Типичный вид графика регрессионных остатков при заниженном порядке аппроксимации.
  - b. Статистические свойства регрессионных остатков.
  - c. Какие из статистических свойств нарушаются при занижении порядка аппроксимации?

- d. Применение критерия серия к анализу коррелированности регрессионных остатков. Какой тип ряда с точки зрения критерия серий формируется при занижении порядка аппроксимации?
  - e. Какую модель следует выбрать, исходя из анализа регрессионных остатков.
3. Скользящий контроль
- a. Что такое эмпирическая модель? Какие еще бывают модели, которые нельзя назвать эмпирическими?
  - b. Что такое обучающая и экзаменационная выборка?
  - c. Какая проблема возникает при оценивании погрешности эмпирической модели по обучающей выборке?
  - d. Что такое проблема переобучения?
  - e. Контроль по одному элементу и контроль по блокам (LOO, q-fold). Расчет дисперсии на скользящем контроле LOO.
  - f. Вывод формул  $\hat{D}_{LOO}$ ,  $RSS_{LOO}$ ,  $TSS_{LOO}$ ,  $R^2_{LOO}$ .
  - g. Проявился ли эффект оптимистической смещенности  $R^2$  относительно  $R^2_{LOO}$ ? Если нет, то какие изменения могли бы к нему привести.
4. Регрессия
- a. Как рассчитывается доверительный интервал прогноза истинного и нового значения при неизвестной дисперсии шума.
  - b. Чем обусловлена погрешность прогноза истинного значения?
  - c. Будет ли погрешность нового значения равна нулю, если  $\hat{\beta} = \beta_{ист}$ ?

Построить поверхность критерия наименьших квадратов  $Q(\beta)$  для своей модели. Построить график изолиний для него же.

#### **Вопросы к защите**

1. Формальная постановка задачи поиска оценок коэффициентов регрессии как задачи минимизации функционала (в векторном виде).
  2. Вероятностная модель единичного элемента выборки, полученного из регрессионной модели.
  3. Модель выборки из регрессионной модели
  4. оценивания неизвестных параметров регрессии по экспериментальным данным.
- 
1. Что такое задача ОМНК, чем она отличается от МНК. С какой целью вводятся веса невязок? Как чисто технически их записать в виде матрицы?
  2. Что такое задача ММНК? Трактовка ММНК как способа учета априорной информации о коэффициентах регрессии. Использование ММНК при обучении нейросетей.
  3. При каких условиях решение задачи ММНК будет совпадать с решением задачи МНК? Проверить предположение, используя формулу аналитического решения задачи ММНК.
  4. Общий вид квадратичной формы. Квадратичная форма как многомерное обобщение квадратного трехчлена. Условие наличия минимума и максимума.