

Практика 12. Оценка погрешности регрессионной модели на скользящем контроле и по доверительному интервалу.

Часть 1. Скользящий контроль

1. Загрузить данные по характеристикам насосных агрегатов по своему варианту.
2. Реализовать скользящий контроль LOO (leave-one-out) и оценить дисперсию шума на скользящем контроле по формуле

$$\hat{D}_{LOO} = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}(x_i))^2$$

3. Оценить дисперсию шума без выделения экзаменационной выборки по формуле

$$\hat{D}_\varepsilon = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}(x_i))^2$$

4. Оценить дисперсию шума без выделения экзаменационной выборки по формуле

$$S_\varepsilon^2 = \frac{1}{n-k} \sum_{i=1}^n (y_i - \hat{y}(x_i))^2$$

5. Сравнить все три дисперсии, сделать выводы (на защите)

Часть 2. Доверительный интервал прогноза

1. Для каждого разбиения выборки на скользящем контроле LOO рассчитать 95%-доверительный интервал прогноза нового значения.
2. Вывести количество и долю попаданий нового значения в доверительный интервал по всем разбиениям. Доля попаданий должна быть близка к 95%.

Вопросы к защите

1. Скользящий контроль
 - a. Что такое эмпирическая модель? Какие еще бывают модели, которые нельзя назвать эмпирическими?
 - b. Что такое обучающая и экзаменационная выборка?
 - c. Эффект оптимистической смещенности при оценивании погрешности эмпирической модели по обучающей выборке. В чем заключается, к чему приводит?
 - d. Что такое проблема переобучения?
 - e. Контроль по одному элементу и контроль по блокам (LOO, q-fold).
 - f. Расчет дисперсии на скользящем контроле LOO. Вывод формул \hat{D}_{LOO} .
 - g. Проявился ли эффект оптимистической смещенности \hat{D}_ε относительно \hat{D}_{LOO} ? К чему
2. Доверительные интервалы регрессии
 - 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 $\hat{\beta} = \beta_{ист}$?