

Практикум 9. Линейная модель. Регрессия.

1. Сначала познакомимся с простой линейной моделью и задачей подгонки прямой под облако точек. Для этого сначала нужно сгенерировать модельные данные: массив X , где $X_k = k/m$ для $m = 10, 20, 50$. Вычислить значения $Y_i = aX_i + b + \varepsilon_i$, где a, b – количество гласных в вашем имени и фамилии соответственно, случайные величины $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ независимы ($\sigma = 1, a/9$).
 - 1.1 Найти оценки для коэффициентов прямой методом МНК (`sklearn.linear_model.LinearRegression`). Построить на одном графике облако точек, полученную прямую, исходную прямую. Сравнить полученные коэффициенты и оценку дисперсии ошибок с исходными.
 - 1.2 * Найти оценки для коэффициентов прямой методом Тейла. Построить на одном графике облако точек, полученную прямую, исходную прямую. Сравнить полученные коэффициенты и оценку дисперсии ошибок с исходными. Построить на одном графике облако точек и прямые, полученные методами Тейла и МНК, сравнить.
2. Теперь посмотрим, как наши методы реагируют на сильно отклоняющиеся от прямой значения Y . Для этого в массиве Y нужно изменить несколько элементов.
 - 2.1 Пусть $Y_t = aX_t + b + 5\sigma$ для некоторого t . Построить новую прямую методом МНК, сравнить с прямой МНК, построенной выше. Если $Y_t = aX_t + b + 5\sigma$, а $Y_s = aX_s + b - 5\sigma$, будут ли эти выбросы компенсировать друг друга?
 - 2.2 * Решить задачу 2.1 для метода Тейла. Сколько значений в массиве Y нужно изменить, чтобы поменялась прямая, построенная методом Тейла?
3. Познакомимся с общей линейной моделью. Посмотрим, как метод МНК работает для подгонки многочлена. Пусть X – тот же массив, $Y_i = a_0 + a_1X_i + \dots + a_kX_i^k\varepsilon_i$, $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ независимы ($\sigma = 1$).
 - 3.1 Построить МНК-оценки коэффициентов, считая что $Y = a_0 + a_1X + \dots + a_dX_i^d + \varepsilon$, где $d = k, k+1, k+3$, сравнить их с исходными. Построить на одном графике облако точек, полученный многочлен, исходный многочлен (отдельно для разных d).
 - 3.2 Построить прогноз в точках $x = 1.2, 1.5, 2$ для разных d , сравнить эти прогнозы. Сравнить графики многочленов, полученных в предыдущей задаче при разных d , за пределами отрезка $[0, 1]$.

Многочлены для задачи 3 ($Y = a_0 + a_1x + \dots + a_kx^k$) по вариантам:

1. $y = x^4 - 3x^2 + x + 1$
2. $y = -2x^4 + 4x^3 - x - 2$
3. $y = x^5 - x^3 - 2x^2 + 3$

Полезные ссылки:

<https://proglab.io/p/linear-regression/>

https://scikit-learn.org/stable/auto_examples/linear_model/plot_ols.html#sphx-glr-auto-examples-linear-model-plot-ols

https://scikit-learn.org/stable/auto_examples/linear_model/plot_ols_ridge_variance.html#sphx-glr-auto-examples-linear-model-plot-ols-ridge-variance

<https://towardsdatascience.com/a-beginners-guide-to-linear-regression-in-python-with-scikit-learn-8>

https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LinearRegression.html