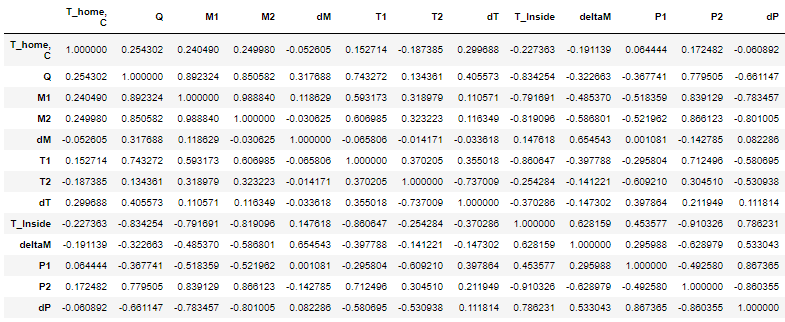
*Предисловие: Я сменил данный с Русской 66 на Давывода 28, Данные взял с 12.19 по 05.20 (полгода), т.к. на портале у меня нет данных о внутренней температуры с Русской. Роман Сергеевич, нужно ли нам использовать зависимости по корреляции?*



*Корреляция данных Давыдова 28.*

*По этой таблице видно, что Q сильно коррелирует от M1, M2, T1 и Внешней температуры. Вносить это наблюдение или оно само собой разумеется и в дипломе будет лишнее?*

1 Часть. Частный случай линейной регрессии. Зависимая Q и независимая Тнв.

Набор данных содержит информацию о тепловых изменениях, записываемых каждый час на контроллере и термометре. Информация включает в себя температуру наружного воздуха, температуру внутри здания, давление в ГВС, тепловая энергия.

Первая задача - предсказать затрачиваемую тепловую энергию для предполагаемого здания, принимая входную функцию за внешнюю температуру.

Q – Отопление

Тнв – температура наружного воздуха

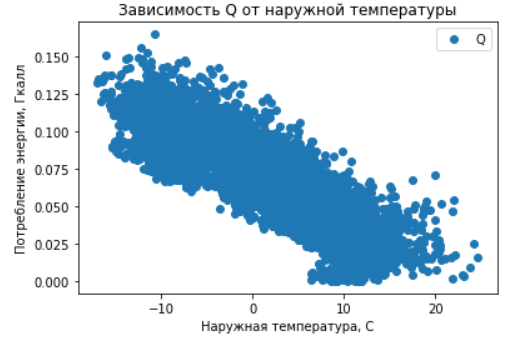
Классическая модель выражается как зависимость интенсивность отопления от температуры наружного воздуха. Представление случая линейной регрессии в виде прямой с одной независимой.

Модель линейной регрессии можно записать следующим образом:

В начале изучим входные данные.

Далее нарисуем наши точки данных на двумерном графике, чтобы увидеть наш набор данных, и посмотрим, сможем ли мы вручную найти какую-либо связь между данными.

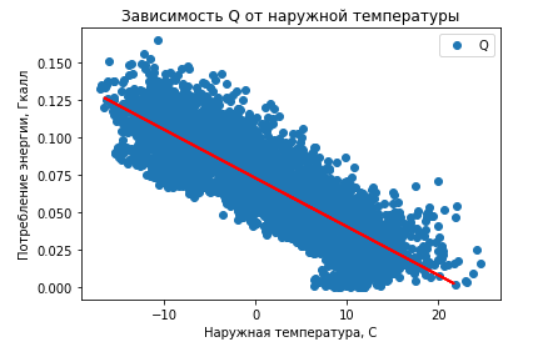
Мы взяли T\_inside и Q для нашего анализа. Ниже представлен двухмерный график между T\_inside и Q.



Атрибуты являются независимыми переменными, в то время как метки являются зависимыми переменными, значения которых должны быть предсказаны. В нашем наборе данных используем только два столбца. Мы хотим предсказать потребление энергии в зависимости от записанногой температуры наружного воздуха. Поэтому наш набор атрибутов будет состоять из столбца «Q», который хранится в переменной X, а метка будет столбцом «T\_inside», который хранится в переменной y.

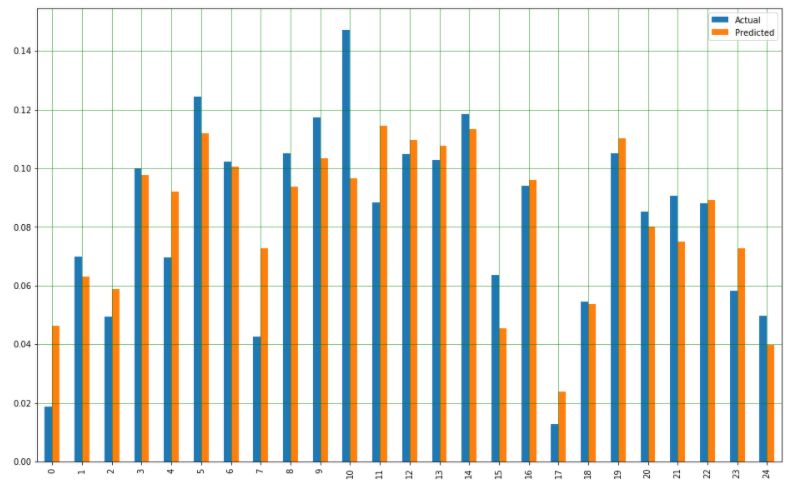
Затем мы разделяем 80% данных на обучающий набор, а 20% данных - на набор тестов.

Значение пересечения и наклона, рассчитанных алгоритмом линейной регрессии для набора данных:

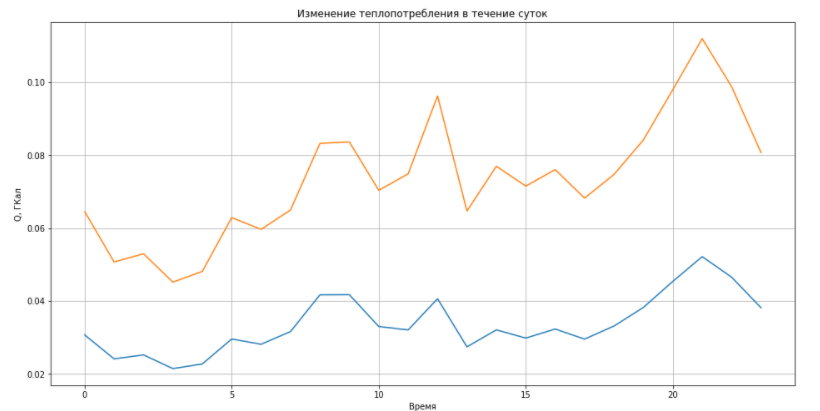


Прогноз против данных испытаний

Прямая линия на графике выше показывает, что наш алгоритм корректен.



Гистограмма, показывающая сравнение фактических и прогнозируемых значений.

Хотя наша модель не очень точна, прогнозируемые проценты близки к фактическим. 

Стандартная модель на основе расчетной нагрузки дает в большинстве случаев завышенное теплопотребление и может быть использована только для контроля перетопов и оценки энергосбережения. Зависимость Q(tнв), построенная на основе фактических данных о потреблении тепловой энергии и измерений температуры наружного воздуха дает более точную аппроксимацию реального потребления тепла, однако ошибка прогноза по такой линии тренда весьма высока, поскольку не учитываются важные параметры теплоносителя и дополнительные погодные условия.