

АНАЛИЗ ВРЕМЕННЫХ РЯДОВ

СОСТАВЛЯЮЩИЕ ВРЕМЕННОГО РЯДА

При анализе временного ряда выделяют три составляющие: тренд, сезонность и шум. Тренд - это общая тенденция, сезонность, как следует из названия - влияния периодичности (день недели, время года и т.д.) и, наконец, шум - это случайные факторы.

Что бы понять отличие этих трёх величин, смоделируем функцию расстояния от земли до луны. Известно, что в среднем луна каждый год отдаляется на 4 см - это тренд, в течение дня луна совершает оборот вокруг земли и расстояние колеблется от ~362600 км до ~405400 км - это сезонность. Шум - это "случайные" факторы, например, влияние других планет. Если мы изобразим сумму этих трёх графиков, то мы получим временной ряд - функцию, показывающую изменение расстояния от земли до луны во времени.

График 1. Тренд

График 2. Сезонность

График 3. Шум

ТРЕНД. МЕТОДЫ СГЛАЖИВАНИЯ

Методы сглаживания необходимы для удаления шума из временного ряда. Существуют различные способы сглаживания, основные - это метод скользящей средней и метод экспоненциального сглаживания.

| МЕТОД СКОЛЬЗЯЩЕЙ СРЕДНЕЙ

Идея метода скользящего среднего заключается в смещении точки графика на среднее значение некоторого интервала. В качестве интервала берут нечётное количество участков, например, три - предыдущий, текущий и следующий периоды, находится среднее и принимается в качестве сглаженного значения:

$$S_i = \sum_{j=-k}^k (x_{i+j}) / (2k+1)$$

У данного метода есть проблема: случайное высокое или низкое значение сильно влияют на скользящую линию. В качестве решения были введены веса. Для распределение веса используют оконные функции, основные оконные функции - это окно Дирихле (прямоугольная функция), В-сплайны, полиномы, синусоидальные и косинусоидальные:

График 4. Окно Ганна для $n=5$ (косинусоидальное окно)

График 5. Синусоидальное окно для $n=5$

Минусы использования скользящей средней - это сложность вычислений и некорректные данные на концах графика.

Исходные данные	Скользящая средняя	Взвешенная скользящая средняя (синусоидальное окно, $n=5$)	Взвешенная скользящая средняя (окно Ганна, $n=5$)
800	840	283	0
879	854	474	400
882	952	517	440
1247	1005	596	441
1012	1038	647	624
1010	1179	652	506
1448	1257	687	505
1556	1216	816	724
848	1343	795	778
1520	1356	756	424
1500	1626	795	760
2637	2094	1110	750
2720	2239	1405	1319
2100	2278	1517	1360
1654	1845	1298	1050
905	1493	945	827
1312	1958	751	453
3962	2469	1188	656
3698	2771	1876	1981
2111	3257	2664	1849

Таблица 1. Сглаживание методом скользящей средней

График 6. Сглаживание скользящей средней. Красный - исходные данные, персиковый - скользящая средняя, жёлтая и охра - скользящая средняя, взвешенная синусоидальным окном с $n=7$ и $n=3$ соответственно

Как видно из графика, увеличение n выдаёт более плавную функцию, таким образом нивелируя более мелкие колебания во временном ряду. Обратите внимание, что при сглаживании не имеет значения, совпадает график среднего с графиком данных или нет, целью является построение **правильной формы**.

| МЕТОД ЭКСПОНЕНЦИАЛЬНОГО СГЛАЖИВАНИЯ

Метод экспоненциального сглаживания получил своё название потому, что в сглаженной функции экспоненциально убывает влияние предыдущего периода с неким коэффициентом чувствительности α . Сглаженное значение находится как разница между предыдущим действительным значением и рассчитанным значением:

$$D'_t = \alpha \cdot D_{t-1} + (1-\alpha) \cdot D'_{t-1}$$

Коэффициент чувствительности, α , выбирается между 0 и 1, в качестве базиса используют значение 0,3. Если есть достаточная выборка, то коэффициент подбирается путём оптимизации.

Исходные данные	Экспоненциальное сглаживание, $\alpha=0,1$	Экспоненциальное сглаживание, $\alpha=0,6$
800	800	800
879	-640	160
882	664	463
1247	-509	344
1012	583	611
1010	-424	363
1448	483	461
1556	-290	684
848	417	660
1520	-291	245
1500	414	814
2637	-223	574
2720	464	1353
2100	-146	1091
1654	341	824
905	-142	663
1312	218	278
3962	-65	676
3698	455	2107
2111	-40	1376

Таблица 2. Экспоненциальное сглаживание

График 7. Экспоненциальное сглаживание с $\alpha=0,1$ (персиковая линия) и $\alpha=0,6$ (жёлтая линия)

МЕТОДЫ ПРОГНОЗИРОВАНИЯ

Методы прогнозирования основываются на выявлении тенденции во временном ряду и последующем использовании найденного значения для предсказания будущих значений. В методах прогнозирования выделяют тренд и сезонность, в общем случае, все типы сезонности могут быть найдены последовательными итерациями. Например, при анализе данных за год, можно выделить сезонность времени года, а в оставшемся тренде найти сезонность по дням недели и так далее.

ДВОЙНОЕ ЭКСПОНЕНЦИАЛЬНОЕ СГЛАЖИВАНИЕ

Двойное экспоненциальное сглаживание выдаёт сглаженное значение уровня и тенденции.

Smooth - сглаживание, сглаженный уровень на период t , s_t , зависит от значения уровня на текущий период (D_t), тренда за предыдущий период (t_{t-1}) и

рассчитанного сглаженного значения на предыдущий период ($s_{\tau-1}$):

$$s_{\tau} = \alpha D_{\tau} + (1 - \alpha)(s_{\tau-1} + t_{\tau-1})$$

Trend - тенденция, тренд на период τ , t_{τ} , зависит от рассчитанного сглаженного значения за предыдущий и текущий периоды (s_{τ} и $s_{\tau-1}$) и от предыдущей тенденции:

$$t_{\tau} = \beta(s_{\tau} - s_{\tau-1}) + (1 - \beta)t_{\tau-1}$$

Рассчитанные по данным формулам уровень и тренд могут быть использованы в прогнозировании:

$$D'_{\tau+h} = s_{\tau} + h \cdot t_{\tau}$$

При расчёте, значения s и t для первого периода назначают $s_1 = D_1$ и $t=0$

График 8. Данные (красная линия), экспоненциальное сглаживание (персиковая линия), тренд (жёлтая линия) и прогноз (линия цвета охры) методом экспоненциального сглаживания

МЕТОД ХОЛЬТ-ВИНТЕРСА

Метод Хольт-Винтерса включает в себя сезонную составляющую, т.е. периодичность. Существуют две разновидности метода - мультипликативный и аддитивный. В отличие от двойного экспоненциального сглаживания, метод Хольт-Винтерса изучает также влияние периодичности.

Общая идея нахождения значений сглаженного уровня, тренда и периодичности заключается в следующем: сглаженный уровень (s - smooth, иногда используют l - level) - это базовый уровень значений, тренд (t - trend) - это показатель скорости роста, разница между сглаженными значениями текущего и предыдущего периода. Для изучения периодичности (p - period), мы разбиваем данные на периоды размером k и выделяем влияние каждого элемента ($1, 2, \dots, k$) периода на сглаженный уровень.

Для более точных расчётов вводится показатель обратной связи. В общем понимании, обратная связь - это влияние предыдущих значений на новые: например, когда Вы начинаете говорить, Вы регулируете громкость своего голоса в зависимости от того, что слышат Ваши уши - это и есть обратная связь.

Для начала расчётов, значения s , t и k , в самом простом виде, могут быть выбраны как $s_\tau = D_\tau$, $t = 0$, $p = 0$.

k - длина выбранного периода:

$$s_\tau = \alpha(D_\tau - p_{\tau-k}) + (1 - \alpha)(s_{\tau-1} + t_{\tau-1})$$

С поправкой на предыдущие значения $t_{\tau-k}$ (обратная связь)

$$t_\tau = \beta(s_\tau - s_{\tau-1}) + (1 - \beta)t_{\tau-1}$$

С поправкой на предыдущие значения $p_{\tau-k}$ (обратная связь)

$$p_\tau = \gamma(D_\tau - s_\tau) + (1 - \gamma)p_{\tau-k}$$

Для прогнозирования используется следующая формула:

$$x_{\tau+h} = D_\tau + ht_\tau + p_{\tau-k+h}$$

МУЛЬТИПЛИКАТИВНЫЙ МЕТОД ХОЛЬТ-ВИНТЕРСА

Мультипликативный метод отличается от аддитивного тем, что параметры, влияющие на периодичность и сглаженный уровень рассчитываются отношением:

$$p_\tau = \gamma(D_\tau/s_\tau) + (1 - \gamma)p_{\tau-k}$$

$$s_\tau = \alpha(D_\tau/p_{\tau-k}) + (1 - \alpha)(s_{\tau-1} + t_{\tau-1})$$

$$t_\tau = \beta(s_\tau - s_{\tau-1}) + (1 - \beta)t_{\tau-1}$$

Для прогнозирования используется следующая формула:

$$x_{\tau+h} = (D_\tau + ht_\tau)p_{\tau-k+h}$$

КАЧЕСТВО ПРОГНОЗИРОВАНИЯ

Проверка качества прогнозирования возможна в случае наличия достаточной выборки и является важной проверкой на достоверность прогноза, для проверки и оптимизации значений α , β и γ необходимо построить прогноз на существующие данные, например, если у нас в наличии данные за пять лет и

мы хотим предсказать следующий год, то необходимо построить модель на первых четырёх годах, проверить и оптимизировать коэффициенты для минимизации ошибки между прогнозом и данными на 5й год. После оптимизации модель может быть перестроена с учётом последнего периода для повышения точности, далее следует построение прогноза.

Методы оптимизации будут описаны в отдельной статье, ниже представлен пример прогнозирования методом Холт Винтерса.

График 9. Данные о посещаемости сайта за четыре недели

#	Данные	s	t	p	s	t	p
1	93	93	0	0	93	0	0
2	91	92	-0.1	-0.5	92	-0.1	0.99
3	72	84	-0.89	-6	84	-0.89	0.93
4	75	80	-1.2	-2.5	80	-1.2	0.97
5	75	77	-1.38	-1	77	-1.38	0.99
6	57	68	-2.14	-5.5	68	-2.14	0.92
7	66	66	-2.13	0	66	-2.13	1
8	123	88	0.28	17.5	38	-4.72	1.62
9	85	87	0.15	-1.25	54	-2.65	1.28
10	85	89	0.34	-5	67	-1.09	1.1
11	91	91	0.51	-1.25	77	0.02	1.08
12	102	96	0.96	2.5	87	1.02	1.08
13	73	90	0.26	-11.25	85	0.72	0.89
14	60	78	-0.97	-9	75	-0.35	0.9
15	99	79	-0.77	18.75	69	-0.92	1.53
16	108	91	0.51	7.88	75	-0.23	1.36
17	98	96	0.96	-1.5	80	0.29	1.16
18	104	100	1.26	1.38	87	0.96	1.14
19	83	93	0.43	-3.75	84	0.56	1.03
20	68	88	-0.11	-15.63	81	0.2	0.86
21	62	81	-0.8	-14	76	-0.32	0.86
22	59	64	-2.42	6.88	61	-1.79	1.25
23	80	66	-1.98	10.94	59	-1.81	1.36
24	121	87	0.32	16.25	76	0.07	1.38
25	112	97	1.29	8.19	85	0.96	1.23
26	85	94	0.86	-6.38	85	0.86	1.02
27	106	106	1.97	-7.82	101	2.37	0.95
28	82	103	1.47	-17.5	100	2.03	0.84

График 9. Пример предсказания посещаемости сайта на основе данных за четыре недели. Жёлтая линия - исходные данные, красная - прогноз на пятую неделю на основе первых четырёх. Закрашена линия сглаженного уровня при $\alpha=0.4$, $\beta=0.1$, $\gamma=0.5$

УДК: ГРНТИ:

Автор статьи: Телятников З.А.

Дата написания статьи: 10.06.2017

Адрес статьи в интернете:

http://k-tree.ru/articles/statistika/prognozirovanie/analiz_vremennih_ryadov

Дата формирования документа: 24.09.2017 18:51

*Все материалы данного файла являются объектами авторского права (в том числе дизайн).
Запрещается копирование, распространение (в том числе путем копирования на другие сайты и ресурсы в Интернете) или любое иное использование информации и объектов без предварительного согласия правообладателя.*