Методичка по временным рядам

github.com/IgorEvsin/ts_course

Игорь Евсин

21 октября 2025 г.

Оглавление

Глава 1

Поехали

https://otexts.com/fpp3/intro.html

1.1 Как мы построили данный курс

Привет! Мы команда увлеченных анализом данных энтузиастов, которых в какой-то момент объединило благородное желание освоить анализ временных рядов. У нас собралось большое количество разнообразных материалов на тему временных рядов из различных областей науки.

Временные ряды - это непросто. Во-многом, дело в том, что основы анализа временных рядов затрагиваются в рамках любой из приведенных ниже дисциплин:

1. Теория вероятностей и математическая статистика
   (a) Во временных рядах есть своя специфика, из-за которой слепое применение классических статистических методов может привести к ошибочным выводам
2. Случайные процессы
   (a) Очень важно, однако сложность изучения не окупается реальными результатами
3. Технические методы временных рядов
   (a) Ранние и во-многом неформализованные способы обработки последовательностей данных
4. Эконометрические методы временных рядов
   (a) Эконометрика, в частности временных рядов, полезна в той части, в которой она описывает особенности данных

(b) Например, нам важно знать такие свойства рядов как стационарность, сезонность, наличие эффекта памяти /автокорреляции, кластеризации волатильности, наличие структурных сдвигов, динамика корреляции между активами. Полезны статистические тесты

(c) Зная эти особенности, мы можем грамотно предобработать данные/подобрать модели В то же время методы предсказания, которые предлагает эконометрика (например, ARIMA/GARCH) практически полностью бесполезны

(d) Следовательно эконометрику следует рассматривать скорее как полезное средство для диагностики, а не решение.

5. Теория обработки сигналов

   (a) Достаточно зрелая и полезная дисциплина, о которой некоторые из нас, в силу своего экономического образования, ничего не знали, за исключением каких-то отрывков знаний вроде фильтра Калмана.

6. Технический анализ

   (a) "Технический анализ - первая попытка человека понять финансовые временные ряды некто со смартлаба.

7. Курсы по машинному обучению / глубокому обучению

   (a) Иногда слепой фит черного ящика сильно вредит. Понимание природы данных усилит надежность/качество моделей.

И это мы еще не говорим уже о сотнях статей в Интернете, посвященных полезным лайфхакам для работы с рядами. Зачастую в них можно почерпнуть действительно полезную информацию, но полной картины они, разумеется, не дают.

Как можно заметить, дисциплин, косвенно касающихся идеи временных рядов довольно много, они разнообразны, при этом между некоторыми областями наблюдается довольно слабое пересечение. Из-за этого полноценное понимание сути временных рядов оказывается затруднено. Поэтому когда мы составляли данный курс, основной антицелью являлось не допустить тупого пересказа имеющейся литературы в этих областях. Вместо этого мы хотели бы, чтобы к его окончанию у студентов развилась интуиция по работе с временными рядами. Мы не претендуем на полноту и точность знаний: курс во многом авторский. Мы также не ставим задачу, чтобы студенты помнили каждую формулу из рассмотренных нами, и не планируем рассматривать каждую модель временного ряда, известную человечеству.

Мы религиозно верим в формулу "знать – значит уметь поэтому горячо рекомендуем прорешать подготовленные нами практические упражнения по временным рядам. Мы очень старались, когда создавали его, так что можете похвалить нас звездочкой на гитхабе.

1.2 Что вообще можно делать с рядами

Вот вы сидите на собеседование и эксцентричный интервьюер вам говорит: "Представь, что у тебя есть временной ряд, что бы ты с ним мог сделать". Это реальный вопрос, который задали на реальном собеседовании одному из авторов данного курса. Поэтому если вы попадете в подобную ситуацию, то не забудьте:

В случае одномерных рядов:

1. Прогнозирование - попытка сделать точное и стабильное предположение о будущем поведении ряда.
2. Декомпозиция (Разложение на компоненты). Фильтрация (Разложение на сигнал и шум).
3. Заполнение пропусков
4. Обработка выбросов, аномалий, детектирование смены режима
5. Дезагрегация - попытка определить поведение ряда внутри рассматриваемых исходных временных интервалов.

С многомерными рядами:

1. То же что и со одним рядом
2. Классификация
3. Кластеризация
4. Поиск взаимосвязей между рядами (расчет корреляций, кросс-регрессий, определение причинности)

Вместе с тем также стоит сказать, что в случае временных рядов одни и те же математические модели могут выполнять несколько функций. Например, модель скользящего среднего может выступать как предиктором ряда, так и его фильтром (срезая его высокочастотные колебания, которые скорее всего являются шумом).

Значительную часть курса мы посвятим техникам прогнозирования одномерного ряда. О том, что такое (хороший) прогноз временного ряда, читайте в следующей главе.

1.3 Генезис временных рядов

Знакомство с анализом временных рядов было бы странно начинать без объяснения, что же собственно временные ряды из себя представляют. К нашему удивлению, авторы многих учебников не удосуживаются этого сделать, а ведь именно в определении скрыта львиная доля критически важных особенностей рядов, без понимания которых их анализ будет ошибочным.

Немного про то, откуда берутся временные ряды

Определение 1.3.1: Случайная функция

Пусть T - произвольное множество, $(\Omega, \mathcal{F}, \mathbb{P})$ - вероятностное пространство.
Тогда отображение

$$X : T \times \Omega \to \mathbb{R}$$

называется случайной функцией, если для любого $t \in T$ функция

$$X(t, \omega) := X_t(\omega) = X_t, \quad \omega \in \Omega,$$

является случайной величиной на $(\Omega, \mathcal{F}, \mathbb{P})$.

Наиболее важные виды случайных функций:

1. случайные процессы (stochastic processes, random processes): $T \subset \mathbb{R}$
   (a) случайные процессы с дискретным временем: $T = \mathbb{Z}_+$ (иногда $T = \mathbb{Z}$)
   (b) случайные процессы с непрерывным временем: $T = [0, \infty)$ (иногда $T = \mathbb{R}$)
2. Случайные поля: $T \subset R^k, \quad k \geq 2$.

Определение 1.3.2: Траектория процесса

Траекторией случайного процесса X_t называется отображение $t \mapsto X_t(\omega)$ при фиксированном ω. Конечномерным распределением случайного процесса X_t называется распределение вектора $(X_{t_1}, X_{t_2}, \ldots, X_{t_n})$ для фиксированного набора моментов времени $t_1, t_2, \ldots, t_n$.

Определение 1.3.3: Временной ряд

Временным рядом Y_t называется набор значений траектории случайного процесса, измеренных в наблюдаемые моменты времени. Он представляет собой дискретизированную или наблюдаемую версию траектории случайного процесса.

Если присмотреться к определению выше, то можно заметить, что главное отличие временного ряда от других типов данных — важность временного порядка. Не только само значение, но и момент его измерения имеют значение. Дело в том, что проводя наблюдения за каким-то природным явлением, мы вовсе не извлекаем получаемые значения из одной и той же генеральной совокупности. Даже если настройки прибора и положение датчиков не менялись, состояние измеряемого объекта в каждый новый момент времени будет другое. Попросту говоря, это будет уже другая случайная величина. Серию измерений, выполненных одним и тем же прибором, даже неподвижно стоящим под одной и той же горой, нельзя рассматривать, как серию выборок из одного и того же пространства элементарных событий. Для описания случайного процесса, в отличие от случайной величины, недостаточно задать его функцию распределения один раз. Просто потому, что в различные моменты времени t она может быть разной. А еще для случайного процесса надо определить функцию совместного распределения вероятностей для моментов времени t и $t + \delta t$ и так далее (чтобы учесть эффект памяти).

Чтобы оценить эти функции, наблюдая за случайным процессом, нужна не одна реализация, а целый ансамбль реализации одного и того же случайного процесса. Тогда и только тогда для каждого момента времени у нас будет несколько измерений одной и той же случайной величины. Только в таком случае становятся применимы все привычные статистические методы. Именно игнорирование данного свойства временных рядов приводят к парадоксальным и ошибочным результатам вроде широко известной статистически значимой корреляции между числом пиратов и глобальным потеплением.

Но что же делать, если у нас есть только одна Земля? Как изучать взаимосвязи между процессами, каждый из которых мы наблюдаем в единственном экземпляре? Проблема эта невероятно сложная, практически нерешаемая. Мы еще вернемся к ней, когда будем говорить о свойствах случайных процессов, в частности о стационарности и эргодичности (см. chapter 4). Пока что укажем, что для некоторых классов случайных процессов, все характеристики которых неизменны во времени, наличие ансамбля не обязательно. То есть, нам не потребуются сотни реализаций, чтобы оценить какую-нибудь статистику, вместо этого достаточно некоторое время понаблюдать за одной. Например, чтобы оценить коэффициент корреляции между X и Y, достаточно иметь одну реализацию X и еще одну – Y.

Но что же делать с остальными рядами, теми которые не имеют таких приятных свойств. Рядами с трендами, сезонными и суточными циклами, и т.д.? Как искать связь между ними и оценивать ее значимость? Как строить предиктивные модели, если даже сходимость оценки автокорреляционной функции не гарантирована? На этот вопрос мы в том числе попытаемся найти ответ в данном учебнике.

Пока же следует запомнить, что общая проблема алгоритмов обработки временных рядов — это отсутствие математической строгости. Ведь даже когда мы используем для анализа экспериментальных сигналов алгоритмы со строгим обоснованием и доказанной оптимальностью, у нас всегда остается открытым вопрос о том, не нарушены ли условия применимости таких алгоритмов? Ведь любой алгоритм всегда начинается с требования, что исходные данные должны обладать вполне определенными свойствами. Но когда мы имеем дело с эмпирическими данными, доказать выполнение этих требований почти невозможно. Поэтому было бы наивно думать, что корректность результата можно гарантировать строгостью метода. На практике использование строгих методов не дает никаких преимуществ, если с тем же уровнем строгости не доказана адекватность модели данных, в рамках которой сформулирован метод. При режимных наблюдениях это почти невозможно. В лучшем случае можно только предполагать, что «базовая модель» сигнала вполне адекватна реальным данным. В худшем (и, к сожалению, более типичном) случае, наоборот, имеются видимые несоответствия между требованиями теоретической модели и экспериментальным сигналом. Но если у нас нет уверенности в адекватности используемой модели данных, это ставит под сомнение и все результаты, полученные в рамках такой модели.

При этом хорошие финальные метрики модели. совершенно не гарантируют, что модель адекватна. Низкая дисперсия остатков на тестовой выборке - это хорошо, однако если это лишь достаточное условия качества модели. Если же свойства временного ряда не соответствуют аксиомам модели, а также условиям проведения эксперимента, то MSE на тесте мало чего говорит о соответствии модели и данных.

Поэтому не надейтесь, что какая-то типовая модель будет хорошо аппроксимировать ваш ряд. В практике изредка встречаются совершенно стандартные временные ряды, в точности подходящие под условия применимости той или иной типовой модели. Но гораздо чаще такого соответствия нет.

С одной стороны, такая постановка проблемы может обескуражить - кажется, что не существует универсального рецепта построения идеальной модели временного ряда. С другой стороны, все это дает право на гибкость и вольность мысли. При изложении материала мы бы хотели, чтобы читатели постоянно держали в голове связь между временным рядом и свойствами породившего его случайного процесса.

Мы верим в то, что именно понимание специфических свойств временного ряда и случайного процесса, его генерирующего, способно привести к пониманию, как построить наиболее точный и устойчивый прогноз. Как следствие, мы не хотим тратить слишком много времени на разъяснение механизмов работы конкретных моделей временных рядов или алгоритмов машинного обучения. Вместо этого мы бы хотели, чтобы у читателей сформировалось понимание, какие особенности исходных данных пытались решить создатели того или иного алгоритма. Сфокусироваться на исходном problem space, а не на текущем solution space (тем более, что он постоянно меняется, появляются все новые алгоритмы предобработки данных, все новые алгоритмы машинного обучения, схемы обучения и т.д.).

После столь пессимистичного начала возникает вопрос: а что же делать и как жить без готовых моделей. Снова оговоримся, что готовых рецептов в данном случае не существует. Однако базовые принципы примерно таковы:

1. Если вы хотите глубоко разобраться в структуре сигнала и научиться его качественно прогнозировать, не пытайтесь сходу применять какие-либо модели к ряду в целом. Постарайтесь сперва понять природу вашего ряда, вникнуть в его экономическую/физическую суть прежде чем приступать к статистическим методам. Это может уберечь вас от многих ошибок при интерпретации результатов количественных методов.

2. Исходя из анализа природы ряда, разберите его на детали. То есть, начните с декомпозиции сигнала на составляющие с максимально простыми свойствами, по возможности опираясь при этом на физику/экономику явления. При проведении данной процедуры постарайтесь пользоваться MECE принципом. MECE принцип кратко описан в section 2.2, наряду с практическими примерами, помогающими освоить этот принцип на практике.

   Далее можно переходить к статистическому разложению.

   > ***Пример :***
   >
   > Например, в экономике это может быть тренд, сезонная и календарная и/или недельная компоненты, эффекты возмущений (праздники и т.п.), разовые события (ковид, военные действия), квазислучайная составляющая и т.д.

   Предварительный предметный анализ, в особенности декомпозиция ряда являются основными элементами разведочного анализа (EDA) временного ряда. Чтобы найти и выделить эти компоненты, начните с разведочного анализа (советую для этих целей ознакомиться с [2])

3. Затем, зная основные элементы сигнала, выделите каждую составляющую в чистом виде. Отдельно - стационарные, отдельно нестационарные. После чего можно строить техническую модель каждой составляющей, разглядывая ее буквально "под микроскопом". Подробнее про составляющие временных рядов мы расскажем в chapter 2

4. Очень многие ряды в принципе не позволяют давать точные прогнозы - таковы внутренние свойства сигнала, в частности рыночные котировки в силу тех или иных гипотез эффективного рынка имеют низкое соотношение сигнал-шум. С такими рядами нужно быть особенно осторожными, поскольку при некорректно сформулированном эксперименте довольно легко подогнать какую-то модель, особенно если у нее достаточно много параметров или если вы переобучитесь на бэктесте. Но как только вы выйдете за тестовые данные, прогноз может сильно потерять в качестве. Подробнее про особенности временных рядов мы расскажем в

5. Поэтому ключевой аспект прогнозирования - это поиск закономерностей в сигнале. Все "модели"фактически именно этим и занимаются. Причем часто их результаты

обусловлены базовыми гипотезами. Предварительный анализ (EDA) по сути для того и нужен, чтобы уловить взаимосвязь между свойствами временного ряда и используемой моделью.

6. Всегда анализируйте остатки. В идеале, они должны быть случайны. Если это не так - значит, модель систематически отклоняется от данных. В лучшем случае это значит, что в сигнале есть какие-то дополнительные закономерности, которые в модели не учтены, и, следовательно, прогноз мог бы быть лучше (если их учесть). В худшем - что модель просто кривая (нарушены условия применимости и т.д.). Уточню еще, что анализировать остатки надо именно по обучающей выборке, а то инет-поиск на "анализ остатков"чаще всего выводит на остатки (погрешности) прогноза, что немного другое). Анализ остатков более подробно мы обсудим в section 1.5.

Источники

1. *adeshere.* Корреляция между временными рядами: что может быть проще? / Хабр. — 16.02.2021. — URL: https://habr.com/ru/articles/542638/ (дата обр. 04.07.2025).
2. *Tukey J.* Разведочный Анализ Данных. — 1981.
3. *Панов В.* Теория случайных процессов. — 2018.

1.4 Что такое (хороший) прогноз

Определение 1.4.1: Определение (из Википедии)

Прогноз (от греч. «предвидение, предсказание») — это научно обоснованное суждение о возможных состояниях объекта в будущем и (или) об альтернативных путях и сроках их осуществления. В узком смысле, это вероятностное суждение о будущем состоянии объекта исследования.

Другими словами, прогноз - это изучение наблюдаемых (реально существующих!) закономерностей и их экстраполяция в будущее. Чудес не бывает. Чем точнее вы выделите и опишете эти закономерности, тем точнее будет прогноз. Чем лучше вы оцените погрешность экстраполяции каждой составляющей, тем более адекватной получится оценка погрешности прогноза в целом.

Очевидно, некоторые вещи прогнозируются лучше, чем другие. Время восхода солнца мы научились предсказывать довольно точно, а вот колебания цен акций все еще остаются для нас загадкой (возможно, навсегда).

На сложность составления прогноза влияют следующие факторы:

1. Как хорошо мы понимаем факторы, влияющие на объект исследования
2. Какое количество данных доступно
3. Насколько будущее похоже на прошлое
4. Влияют ли наши прогнозы на поведение объекта, которое мы пытаемся спрогнозировать (частный случай - самоисполняющееся пророчество)

TODO: What can be forecast? Взять 4 пункта и написать, что делать, если они не выполняются. Для judgmental forecasting дать ссылку на суперфоркастинг и, возможно, на темы, которые там не покрываются (если такие есть). Футурология.

Статистическую модель же (в самом общем понимании этого слова) можно считать хорошей при соблюдении двух условий:

1. Во-первых, в остатках не должно быть явных закономерностей. Попросту говоря, остатки должны быть случайны. Это значит, что наша модель уже учла все, что можно (нужно?). В модели не должно быть смещения по какой-то из фичей, а смещение по времени должно иметь объяснимый характер.
2. Хороший прогноз - это на самом деле НЕ точный прогноз (как многие ошибочно думают), а прогноз с достоверно известной погрешностью, причем очень желательно - минимально возможной для данного ряда.

Последнее условие очень важно. На протяжении всего курса мы будем вас предостерегать не полагаться исключительно на статистические критерии (в случае временных рядов они будут очень часто лгать). Дело в том, что хорошее качество на тестовой выборке можно получить очень нечестными путями, например через переобучение на тестовой выборке.

Best worst corner case

Именно поэтому мы убеждены, что количество параметров модели должно быть минимально возможным (но не меньше!). По крайней мере, оно должно быть кратно меньше, чем число степеней свободы данных. Иначе возникает риск переподгонки, что ничуть не менее опасно, чем неправильная модель. Правильная оценка числа степеней свободы особенно важна, если в данных есть внутренние взаимозависимости. Ведь если они сильны, то реальное число степеней свободы может быть много меньше, чем формальное количество значений данных.

Пример :

Кстати, известный баг с ложными корреляциями - это типичный пример именно такой ситуации (значения временного ряда взаимозависимы, поэтому фактическое число степеней свободы на порядки меньше, чем количество точек данных).

В качестве награды за большое количество прочитанного материала наградим вас вашей первой моделью временного ряда

Определение 1.4.2: Наивный прогноз

Для наивного прогноза мы берем предыдущее значение ряда в качества прогноза для следующего шага.

$$\hat{y}_{T+h|T} = y_T.$$

Пример : Наивный прогноз

Не смейтесь над простотой этого метода.

Во-первых, он служит хорошим бейзлайном для понимания метрик более сложных моделей. Делать выводы на основании MSE или любой другой выбранной метрики (подробнее в следующей секции) на бэктесте довольно сложно без какого-то референсного значения.

Во-вторых, его можно посчитать для любого ряда и очень быстро.

Во-третьих, когда вы будете работать с реальными данными, вы удивитесь, насколько в некоторых случаях будет незначительна разница между прогнозами сложных моделей и подобным наивным методом.

Наивный метод хорошо работает для временных рядов, порожденных мартингалами.

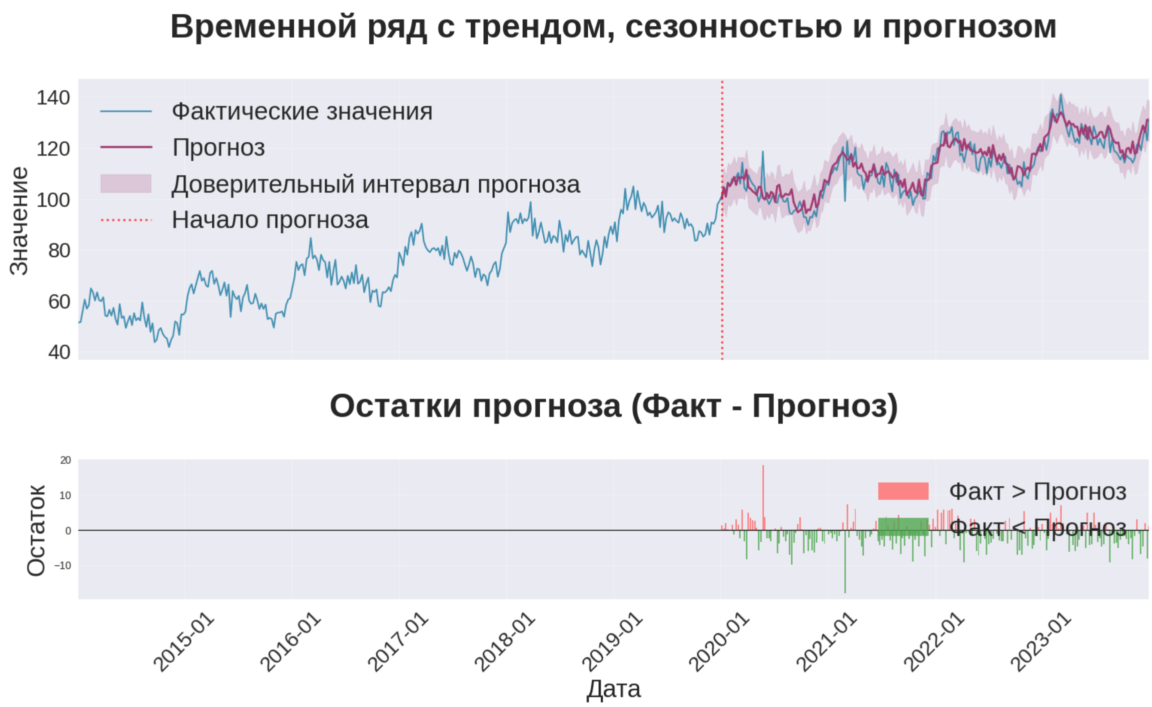
Можно ли придумать модель с меньшим числом свобод, чем наивный прогноз? Вот то-то.

1.5 Метрики качества точечного прогноза

1.5.1 Немного определений

В контексте всего вышесказанного определимся, как же количественно посчитать, насколько наш прогноз совпадает с фактом. Мы уже выяснили, что важны не только конкретные значения метрик, но и их стабильность во времени. На время пока абстрагируемся от этой идеи и напомним, что же метрики качества есть такое.

Допустим, мы построили некоторый предиктивный алгоритм. Мы можем применить его к нашему ряду на некоторой тестовой выборке.



В правой нижней части графика отображены ошибки прогноза в виде барчарта. Анализ остатков, критически важная вещь в классическом машинном обучении / эконометрике, в случае временных рядов играет еще большую роль, ведь в этом случае необходимо также проверять, что в ошибках нет смещения по времени измерения. Например, в нашем случае в остатках есть четкая сезонная составляющая - присутствие этого паттерна говорит нам, что наша предиктивная модель была построена не лучшим образом. Мы хотим, чтобы остатки были случайны, чтобы в них не наблюдались никакие ярко выраженные паттерны.

Теперь перейдем к пониманию, зачем нужны метрики качества. Мы видим, что ошибки прогноза на тестовой выборке отличаются: какие-то из них отрицательные, какие-то положительные, какие-то имеют высокоамлитудные, какие-то близки к нулю. Конечно, мы бы хотели, чтобы все остатки прогноза были близки к нулю. На практике однако подобное недостижимо, поэтому нам нужен некоторая функция агрегация, которая бы значения всех этих остатков превратила в одно число. Тогда мы могли бы сравнивать различные модели по тому, насколько это число больше или меньше. Данная функция как правило задается двумя компонентами: выбором функции потерь для индивидуального прогноза и способа их агрегации. Выбор способа агрегации влияет на то, как сильно мы ценим то или иное отклонение факта от прогноза. Скажем, мы можем ценить негативное отклонение факта от прогноза сильнее, чем позитивное - в таком случае мы будем использовать что-то вроде квантильной функции потерь. Мы также можем ценить большие отклонения сильнее чем небольшие - в таких случаях мы скорее всего будем использовать что-то вроде MSE вместо MAE. От способа агрегации функций потерь зависит то, какие наблюдения с какими отклонениями мы ценим выше. Скажем, если в качестве способа агрегации мы выбрали среднее, мы ценим все наблюдения в тестовой выборке одинаково. Если же мы выберем в качестве метода агрегации, например, 95% квантиль распределения, то мы будем ценить больше наблюдения с высокоамлитудными отклонениями.

Если говорить кратко, метрики качества задают то, насколько мы ценим те или иные ошибки. Исходя из бизнес-цели, мы можем подобрать ту или иную метрику качества. По сути различие между конкретными функционалами качества заключается в том, какие ошибки считаются более, а какие менее приоритетными.

Можно также сказать, что метрики качества - это дескриптивные статистики распределения остатков.

Пример : Общеизвестный пример

Про прогноз запасов товаров на складе

Закономерно вытекает определение выбросов модели

Определение 1.5.1: Выбросы (или аномалии)

Выбросы - это outliers гистограммы распределения остатков

Когда выбросы определяются на основании гистограмм исходных данных, с последующим применением чего-то вроде правила трех сигм, в качестве прогнозной модели неявно предполагается константная модель.

метрики качества моделей временных рядов как некоторые дескриптивные статистики распределения остатков модели. аномалии сюда же (выбросы распределений). Стьюдентизированные остатки?

https://mlgu.ru/3091/

Чекни также Бабушкина, раздел про метрики, там про ряды как раз

TODO: добавить ссылки на вшэ, гевиссту, сайт выше

TODO: написать про свойства метрик качества. типа интерпретируемости, симметричности, устойчивости к выбросам и тд и в конце привести сводную таблицу. Выделить классы эквивалентности

1.5.2 Самые распространенные метрики качества прогноза

Основная задача, с которой мы будем сталкиваться в данным курсе - задача регрессии. Способов посчитать близость двух чисел (прогноза и истинного ответа) достаточно много, и поэтому при обсуждении регрессии у нас возникнет большое количество функционалов ошибки. Напомним также, что в случае регрессии (в частности, в прогнозе временных рядов) функционал и метрика качества зачастую оказывается одним и тем же: мы можем рассматривать среднеквадратичную ошибку как в качестве цели для оптимизации, так и в качестве оценки итогового качества модели. При этом не все метрики качества окажутся легко интерпретируемыми.

Чтобы обучать регрессионные модели временного ряда, нужно определиться, как именно измеряется качество предсказаний. Будем обозначать через y значение целевой переменной, через $\hat{y}_t$ – прогноз модели. Рассмотрим несколько способов оценить отклонение $L(y, \hat{y}_t)$ прогноза от истинного ответа.

MSE и R^2

Основной способ измерить отклонение — посчитать квадрат разности:

$$L = (\hat{y}_t - y)^2$$

Благодаря своей дифференцируемости эта функция наиболее часто используется в задачах регрессии. Основанный на ней функционал называется среднеквадратичным отклонением (mean squared error, MSE):

$$\text{MSE} = \frac{1}{T}\sum_{t=1}^{T}(\hat{y}_t - y_t)^2$$

Отметим, что величина среднеквадратичного отклонения плохо интерпретируется, поскольку не сохраняет единицы измерения — так, если мы предсказываем цену в рублях, то MSE будет измеряться в квадратах рублей. Чтобы избежать этого, используют корень из среднеквадратичной ошибки (root mean squared error, RMSE):

$$RMSE = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(\hat{y}_t - y_t)^2}.$$

Среднеквадратичная ошибка подходит для сравнения двух моделей или для контроля качества во время обучения, но не позволяет сделать выводы о том, насколько хорошо данная модель решает задачу. Например, MSE = 10 является очень плохим показателем, если целевая переменная принимает значения от 0 до 1, и очень хорошим, если целевая переменная лежит в интервале (10000, 100000). В таких ситуациях вместо среднеквадратичной ошибки полезно использовать коэффициент детерминации (или коэффициент R^2):

$$R^2 = 1 - \frac{\sum_{t=1}^{T}(\hat{y}_t - y_t)^2}{\sum_{i=t}^{T}(y_t - \bar{y})^2},$$

где $\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$ — среднее значение целевой переменной. Коэффициент детерминации измеряет долю дисперсии, объяснённую моделью, в общей дисперсии целевой переменной. Фактически, данная мера качества — это нормированная среднеквадратичная ошибка. Если она близка к единице, то модель хорошо объясняет данные, если же она близка к нулю, то прогнозы сопоставимы по качеству с константным предсказанием.

MAE

Заменим квадрат отклонения на модуль:

$$L = |\hat{y}_t - y|$$

Соответствующий функционал называется средним абсолютным отклонением (mean absolute error, MAE):

$$MAE = \frac{1}{T}\sum_{t=1}^{T}|\hat{y}_t - y_t|.$$

Модуль отклонения не является дифференцируемым, но при этом менее чувствителен к выбросам. Квадрат отклонения, по сути, делает особый акцент на объектах с сильной ошибкой, и метод обучения будет в первую очередь стараться уменьшить отклонения на таких объектах. Если же эти объекты являются выбросами (то есть значение целевой переменной на них либо ошибочно, либо относится к другому распределению и должно быть проигнорировано), то такая расстановка акцентов приведёт к плохому качеству модели. Модуль отклонения в этом смысле гораздо более терпим к сильным ошибкам.

MAPE и SMAPE.

В задачах прогнозирования нередко измеряется относительная ошибка. Во-первых, это удобно для интерпретации — легко понять, что «ошибка 50%» соответствует отклонению в полтора раза от целевой переменной. Во-вторых, это позволяет работать с разными масштабами. Например, мы можем решать задачу прогнозирования спроса на товары в магазине, и какие-то товары могут продаваться штуками, а какие-то — тысячами. Чтобы при усреднении ошибок более популярные товары не оказывали большее влияние на результат, следует использовать функции потерь, не зависящие от масштаба. Типичный пример относительной функции потерь:

$$L = \frac{|y_t - \hat{y}_t|}{y_t}$$

Соответствующий функционал называется средней абсолютной процентной ошибкой (mean absolute percentage error, MAPE).

У MAPE есть проблема с несимметричностью: скажем, если $y = 1$ и все прогнозы неотрицательные, то максимальная ошибка при занижении прогноза ($a < y$) равна единице, а ошибка при завышении прогноза ($a > y$) никак не ограничена сверху. Это исправляется в симметричной модификации (symmetric mean absolute percentage error, SMAPE):

$$L = \frac{|y - \hat{y}_t|}{(|y| + |\hat{y}_t|)/2}$$

TODO: квантильная функция

1.5.3 Для тех, кому все еще мало

Как уже было сказано ранее, метрик качеств для временных рядов можно придумать бесконечное количество. Ниже мы приведем менее распространенные метрики качества прогноза, а также дадим краткое объяснение, зачем они нужны.

TODO: написать также про следующие метрики (взято у гевиссты, их дохуя, поэтому надо расписать, для чего вообще они нужны + возможно построить как выглядят график их функции):

1. Метрики качества, которые зависят от масштаба данных (RMSE, MSE, MAE, MdAE, RMSLE, MSLE)
2. Корень из среднеквадратичной ошибки (root mean squared error, RMSE)
3. Средняя абсолютная ошибка (mean absolute error, MAE)
4. Медианная абсолютная ошибка (median absolute error, MdAE)
5. Сравнение RMSE, MAE, MdAE
6. Корень из среднеквадратичной логарифмической ошибки (root mean squared logarithmic error, RMSLE)
7. Метрики качества на основе процентных ошибок (MAPE, MdAPE, sMAPE, sMdAPE, WAPE, WMAPE, RMSPE, RMdSPE)
8. Средняя абсолютная процентная ошибка (mean absolute percentage error, MAPE)
9. Медианная абсолютная процентная ошибка (median absolute percentage error, MdAPE)
10. Симметричная средняя абсолютная процентная ошибка (symmetric mean absolute percentage error, SMAPE)

11. Симметричная медианная абсолютная процентная ошибка (symmetric median absolute percentage error, SMdAPE)

12. Взвешенная абсолютная процентная ошибка (weighted absolute percentage error, WAPE)

13. Средневзвешенная абсолютная процентная ошибка (weighted mean absolute percentage error, WMAPE)

14. Корень из среднеквадратичной процентной ошибки (root mean square percentage error, RMSPE)

15. Корень из медианной квадратичной процентной ошибки (root median square percentage error, RMdSPE)

16. Сравнение MAPE, MdAPE, sMAPE, sMdAPE, WAPE, WMAPE, RMSPE, RMdSPE

17. Метрики качества на основе относительных ошибок (MRAE, MdRAE, GMRAE)

18. Средняя относительная абсолютная ошибка (mean relative absolute error, MRAE)

19. Медианная относительная абсолютная ошибка (median relative absolute error, MdRAE)

20. Средняя геометрическая относительная абсолютная ошибка (geometric mean relative absolute error, GMRAE)

21. Сравнение MRAE, MdRAE, GMRAEV.4. Относительные метрики качества (RelMAE, RelRMSE)

22. Относительная метрика MAE (relative MAE, RelMAE)

23. Относительная метрика RMSE (relative RMSE, RelRMSE)

24. Метрики качества на основе масштабированных ошибок (MASE)

25. Средняя абсолютная масштабированная ошибка (mean absolute scaled error, MASE)

1.5.4 Сравнение метрик качества

Сравнение метрик качества имеет смысл только если оно проводится статистически! То есть с учетом распределений данных, их размера и числа попыток тестирования. Критерий Диболда-Мариано

Вопросы для закрепления

Пример : Вопросы

1. Вопрос 1
2. Вопрос 2
3. Вопрос 3

Глава 2

Компоненты временных рядов

2.1 Мотивация

Зачем это нужно, про консалтеров немного и ссылки

2.2 MECE декомпозиция

Пример :
Если перед вами стоит задача предсказания прибыли компании, имеет смысл предварительно представить ее в виде разности выручки и затрат и далее отдельно прогнозировать каждую компоненту. Вспомним базовую микроэкономику:

$$\begin{aligned}
\text{Profit} &= \text{Total Revenue} - \text{Total Costs} \\
\text{Total Revenue} &= \text{Average Check} \times \text{Quantity Sold} \\
\text{Total Costs} &= \text{Variable Costs} + \text{Fixed Costs} \\
\text{Variable Costs} &= \text{Average Variable Cost} \times \text{Quantity Sold} \\
\text{Profit} &= (\text{Average Check} - \text{Average Variable Cost}) \times \text{Quantity Sold} - \text{Fixed Costs}
\end{aligned}$$

Разложив прибыль в итоговом виде, мы можем создать предсказательную модель для каждого элемента и затем получить итоговый прогноз прибыль, собрав в композицию прогнозы.

Глава 3

Технические особенности временных рядов

Определение 3.0.1: Эквидистантность

Эквидистантный временной ряд – это последовательность измерений, сделанных через равные промежутки времени.

Пример :

На практике интервалы между измерениями часто бывают неравными, особенно в экономике или при наблюдении за природными явлениями. Например, данные могут собираться только в рабочие дни, что приводит к пропускам на выходные и праздники. Эти пропуски нельзя игнорировать, особенно если временные лаги играют важную роль (например, при доставке скоропортящихся товаров). Такие увеличенные интервалы не всегда можно просто выбросить, особенно если речь идет о непрерывных производствах или фиксированных временных лагах. Например, если для производства товара А нужен скоропортящийся товар Б, на доставку которого надо 3 дня, то рост производства Б в понедельник и вторник отобразится в товаре А на третий рабочий день, а рост производства Б в пятницу - уже в понедельник (а это следующая точка, если выходные отброшены).

Глава 4

Свойства случайных процессов

Часть III

Свойства случайных процессов
Основано на [1], [3]

4.1 Стационарность

Определение 4.1.1: Стационарность в узком смысле

Случайный процесс X_t называется стационарным (stationary, стационарным в узком смысле), если все его конечномерные распределения инвариантны относительно сдвигов, т.е. для любых наборов моментов времени $t1, \ldots, tn$, любых вещественных $x1, \ldots, xn$ и любого $h > 0$:

$$\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) = \mathbb{P}(X_{t_1+h} \leq x_1, \ldots, X_{t_n+h} \leq x_n),$$

Определение 4.1.2: Стационарность в широком смысле

Случайный процесс X_t называется стационарным в широком смысле (wide sense stationary, weakly stationary, covariance stationary, second-order stationary), если $m(t)$ является постоянной величиной (не зависящей от t), и кроме того, для любых $h > 0$, s, $t \in \mathbb{R}$ выполнено

$$K(t+h, s+h) = K(t, s).$$

Другими словами, процесс с постоянным математическим ожиданием является стационарным в широком смысле тогда и только тогда, когда существует функция $\gamma : \mathbb{R} \to \mathbb{R}$ такая, что $K(t, s) = \gamma(t - s)$ для любых t, s.

В контексте анализа временных рядов нам обычно достаточно выполнение стационарности в широком смысле.

Пример : Виды стационарностей

1) Узкая стационарность: Все данные на все точки строго из одного распределения.

2) Широкая стационарность: Первая половина элементов взята из экспоненциального распределения, а вторая половина из нормального. Но 1 и 2 моменты у них совпадают и автоковариация зависит только от лага => ряд стационарен в широком смысле ■

Функция $\gamma(\cdot)$ называется автоковариационной функцией (autocovariance function) и обладает следующими свойствами.

Предпосылка 4.1.3: Свойства автоковариационной функции

Пусть γ — автоковариационная функция некоторого стационарного в широком смысле случайного процесса. Тогда

(i) $\gamma(0) \geq 0$.

(ii) $|\gamma(u)| \leq \gamma(0)$.

(iii) γ является четной функцией.

4.2 Эргодичность

Понятие эргодичности мотивировано законом больших чисел. Рассмотрим процесс X_t, наблюдаемый в дискретные моменты времени $t = 1, 2, \ldots, T$ и зададимся вопросом, сходится ли процесс

$$M_T = \frac{1}{T} \sum_{t=1}^{T} X_t$$

при устремлении горизонта времени $T \to \infty$.

Определение 4.2.1

Процесс X_t с дискретными временами $t = 1, 2, \ldots$ называется *эргодическим по среднему*, если

$$M_T \xrightarrow{p} \mu, \quad \text{при } T \to \infty$$

где μ — некоторая константа.

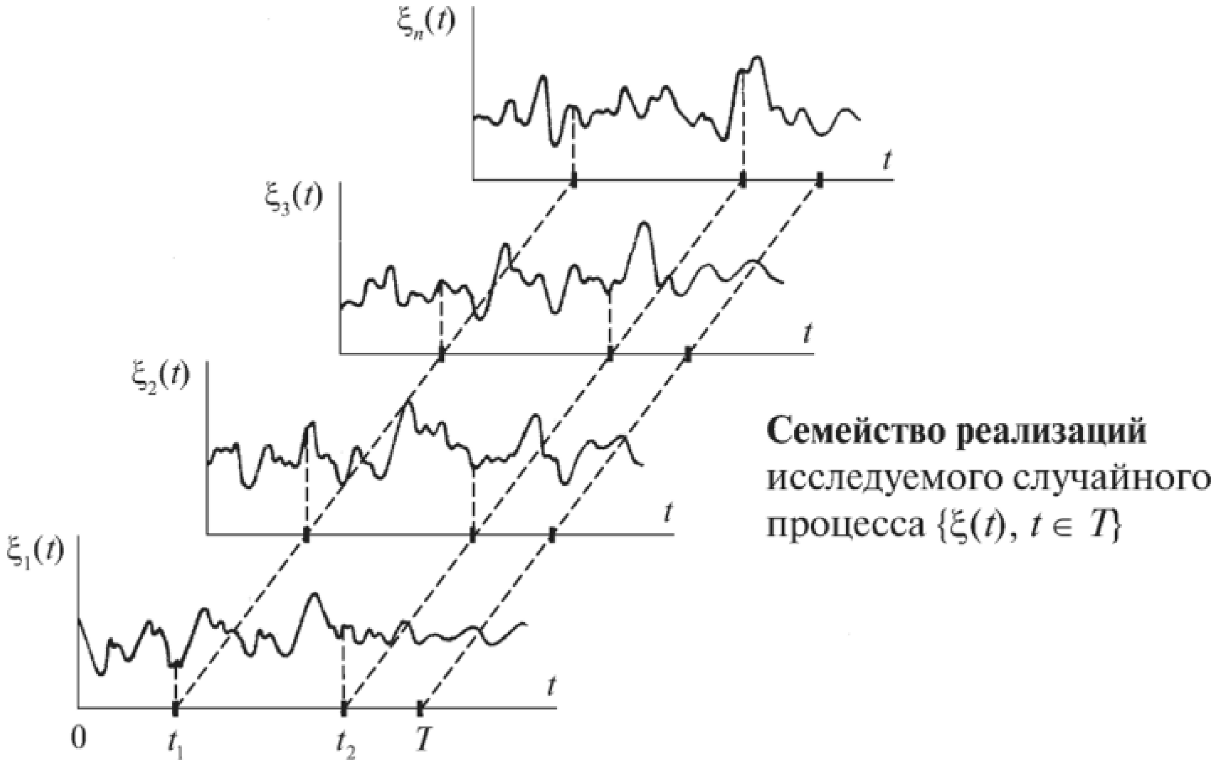


Рис. 4.1: Семейство реализаций СП

Давайте вернемся к тому, с чего начинали. Во введении мы говорили о необходимости иметь множество реализаций случайного процесса для того, чтобы иметь возможность адекватно посчитать по ним статистики. Оказывается, что для эргодических случайных процессов, все характеристики которых неизменны во времени, наличие ансамбля не обязательно! То есть, нам не потребуется десять реализаций, чтобы оценить какую-нибудь статистику. Вместо этого достаточно некоторое время понаблюдать за одной! Например, чтобы оценить коэффициент корреляции между X и Y, достаточно иметь одну реализацию X и еще одну – Y. Что, собственно, все мы и делаем, когда вычисляем коэффициент корреляции между потеплением и пиратами.

При анализе наблюдений очень часто априори считается, что исследуемые процессы являются эргодическими. Иногда это даже не оговаривается специально. Но если мы хотим избежать грубейших ошибок, то нельзя забывать, что гипотеза эргодичности – это только гипотеза.

Подавляющее большинство долговременных наблюдений продолжается конечное время (вы поняли, это такая шутка), а на выходе получается единственный ряд. Доказать эргодичность такого процесса в принципе невозможно. Поэтому, начиная анализ данных, мы чаще всего просто постулируем ее явным образом или неявно. А что еще остается делать, если в наличии куча данных и руки чешутся начальник требует срочно использовать всю мощь безупречного, многократно проверенного теоретиками статистического инструментария для достижения практических целей?

На самом деле, подавляющее большинство временных рядов вовсе не являются эргодическими. И если доказать эргодичность процесса достаточно сложно (практически нереально), то вот опровергнуть ее часто можно без особых усилий. Достаточно просто вспомнить, что практически все экспериментальные временные ряды существенно нестационарны. Огромный массив накопленных экспериментальных данных однозначно свидетельствует, что априорная "базовая модель"почти любого природного процесса – это вовсе не белый шум (для которого действительно можно заигрывать с эргодичностью). Нет, спектры большинства реальных сигналов имеют степенной вид (Точнее, они обычно становятся степенными после удаления доминирующих периодичностей - сезонной, суточной и т.д.)

Но если ряд не стационарен, то он заведомо не может рассматриваться, как последовательность измерений одной и той же случайной величины. Для него совершенно бессмысленно оценивать те статистики, которые вводятся и исследуются при анализе случайных величин.

Из стационарности не следует эргодичность.

Пример : Рассммотрим некоторый случайный процесс и проверим его свойства

Пусть $X(t)$ — белый шум со случайной мощностью, заданный в виде:

$$X(t) = \sigma \cdot W(t),$$

где:

- $W(t)$ — стандартный белый шум в дискретном времени $t \in \mathbb{Z}$, $E[W(t)] = 0$, единичной дисперсией $\mathrm{Var}[W(t)] = 1$, и автокорреляцией $R_W(\tau) = \delta(\tau)$,

  [a]

- σ — случайная величина, не зависящая от $W(t)$, с $E[\sigma^2] < \infty$.

1.Сначала докажем стационарность процесса:

Для этого вспомним условие стационарности в широком смысле:

- Среднее не зависит от времени:

$$E[X(t)] = E[\sigma \cdot W(t)] = E[\sigma] \cdot E[W(t)] = 0.$$

- Автокорреляционная функция зависит только от разности $\tau = t_1 - t_2$:

$$R_X(t_1, t_2) = E[X(t_1)X(t_2)] = E[\sigma^2 W(t_1)W(t_2)] = E[\sigma^2] \cdot \delta(t_1 - t_2).$$

Обозначим $R_X(\tau) = E[\sigma^2]\delta(\tau)$, что зависит только от τ.

Вывод:

Процесс $X(t)$ **стационарен**, так как его среднее постоянно, а автокорреляция зависит только от временного сдвига.

2. Далее докажем его неэргодичность:

2.1 Среднее значение:

- По ансамблю: $E[X(t)] = 0$ (доказано выше).
- По времени (для одной реализации):

$$\lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} X(t)\,dt = \sigma \cdot \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} W(t)\,dt = 0 \quad \text{(т.к. } W(t) \text{ имеет нулевое среднее).}$$

Здесь совпадение есть.

2.2 Дисперсия (второй момент):

- По ансамблю:

$$E[X^2(t)] = E[\sigma^2 W^2(t)] = E[\sigma^2] \cdot E[W^2(t)] = E[\sigma^2] \cdot \delta(0).$$

(Для дискретного белого шума $\delta(0) = 1$, для непрерывного — обобщённая функция.)

- По времени (для одной реализации):

$$\lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} X^2(t)\,dt = \sigma^2 \cdot \lim_{T\to\infty} \frac{1}{T} \int_{-T/2}^{T/2} W^2(t)\,dt = \sigma^2 \cdot E[W^2(t)] = \sigma^2.$$

Но σ^2 — случайная величина, а не константа $E[\sigma^2]$.

Вывод:

- Для **среднего** эргодичность есть.
- Для **дисперсии (и более высоких моментов)** эргодичности **нет**, так как усреднение по времени даёт σ^2, а не $E[\sigma^2]$.

■

[a] Функция Дирака $\delta(\tau)$ определяется как $\delta(0) = 1$ и $\delta(\tau) = 0$ при $\tau \neq 0$ в дискретном случае; в непрерывном времени это обобщённая функция (дистрибуция) с $\int \delta(\tau) d\tau = 1$.

Контрольные вопросы : Свойства случайных процессов

1. Что такое стационарность в узком и широком смысле?

2. Зачем нужна стационарность? Какие из алгоритмов машинного обучения требуют стационарности исходных данных?

3. Как визуально на основании графика определить, является ли процесс стационарным

Источники

1. *adeshere*. Корреляция между временными рядами: что может быть проще? / Хабр. — 16.02.2021. — URL: https://habr.com/ru/articles/542638/ (дата обр. 04.07.2025).
2. *Tukey J.* Разведочный Анализ Данных. — 1981.
3. *Панов В.* Теория случайных процессов. — 2018.