Методичка по временным рядам

github.com/IgorEvsin/ts_course

Игорь Евсин

8 июля 2025 г.

Оглавление

Глава 1

Поехали

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

1.1 Как мы построили данный курс

У нас собралось большое количество разнообразных материалов на тему временных рядов из различных областей науки. Они все довольно разнообразны, при этом между областями слабое пересечение. Из-за этого полноценное понимание сути временных рядов не удаётся. Мы поставили задачу связать все эти области воедино. Мы не претендуем на полноту и точность знаний: курс во многом авторский. Мы хотим, чтобы к его окончанию у студентов развилась интуиция в плане работы с временными рядами. Мы не ставим задачу, чтобы студенты помнили каждую формулу из рассмотренных нами, и не планируем рассматривать каждую модель временного ряда, известную человечеству.

Дело в том, что временные ряды включают элементы следующих систем:

1. ТВиМС
2. Случайные процессы
3. Технические методы временных рядов
4. Эконометрические методы временных рядов
5. Обработка сигналов
6. Технический анализ
7. ML
8. DL

Стратегия: надо накидать структуру курса и затем включать в неё все имеющиеся материалы. Если что-то не вмещается в структуру, надо обновить структуру.

Тактика: выложить на GitLab методичку и лекции. Идти последовательно: я и ассистенты создаём лекции по одной. Когда заканчиваем одну, делаем другую. Пишем ноутбуки и сами же разрабатываем домашние задания (не слишком усложняя их).

Формат курса:

1. Методичка. Источником вдохновения является методичка по ПЛА. Со ссылками на дополнительные источники: всё равно всё покрыть невозможно, главное — дать базу.

2. Ноутбуки. 16 ноутбуков, соответствующих 15 коммодитис из индекса. Цель каждого задания — спрогнозировать ряд с наилучшей метрикой качества, используя изученные методы.

3. Очень хочется обойтись без слайдов.

1.2 Что стоит включить

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.

Описать, с чем мы будем работать — случаи, когда данные всё-таки есть.

Чек-лист построения (ML) модели. Взять за основу разделы 1.3, 1.4, 1.6. Зачем нужны чек-листы? Не для механического построения моделей, а чтобы не забывать о важных шагах.

Связь случайных процессов с временными рядами. Глава 1.7, но более математизированная. Возможно, упомянуть факты вроде того, что условное матожидание — лучший прогноз для МСК.

Мартингал. Наивный прогноз. Задание: построить наивный прогноз для простого ряда. Можно ненавязчиво включать разные официальные источники данных в курс.

Основано на [1], [2]

Определение 1.1

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 1.2

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

Определение 1.3

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.

Эквидистантный временной ряд — это последовательность измерений, сделанных через равные промежутки времени. Однако 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).

Главное отличие временного ряда от других типов данных — важность временного порядка. Не только само значение, но и момент его измерения имеют значение.

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 разные моменты времени t она может быть разной. А еще для случайного процесса надо определить функцию совместного распределения вероятностей для моментов времени t и t+dt и так далее.

Чтобы оценить эти функции, наблюдая за случайным процессом, нужна не одна реализация, а целый ансамбль. Ну, хотя бы десяток реализаций. Причем, это обязательно должны быть 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. Как их обрабатывать дальше, мы уже знаем из школьного вузовского курса статистики.

Это часто игнорируется при применении стандартных статистических методов, что может привести к ошибочным выводам, например, к ложной 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. Оказывается, что для некоторых классов 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.

Но что же делать с остальными рядами, теми которые не имеют данных приятных свойств. Рядами с трендами, сезонными и суточными циклами, и т.д.? Как искать связь между ними и оценивать ее значимость? 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 преамбулы (требования), что исходные данные должны обладать вполне определенными свойствами. Но когда мы имеем дело с экспериментальным сигналом, 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.

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 способно привести к пониманию, как построить наиболее точный и устойчивый прогноз. Как следствие, мы не хотим делать слишком большой упор 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.).

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. *Панов В.* Теория случайных процессов. — 2018.

1.3 Измерение ошибки в задачах точечного предсказания временного ряда

https://mlgu.ru/3091/

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

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 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 a – прогноз модели. Рассмотрим несколько способов оценить отклонение $L(y, a)$ прогноза от истинного ответа.

MSE и R^2

Основной способ измерить отклонение — посчитать квадрат разности:

$$L(y, a) = (a - y)^2$$

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

$$\text{MSE}(a, X) = \frac{1}{T}\sum_{t=1}^{T}(a_t - y_t)^2$$

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):

$$RMSE(a, X) = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(a_t - y_t)^2}.$$

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):

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

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:

$$L(y, a) = |a - y|$$

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

$$MAE(a, X) = \frac{1}{T}\sum_{t=1}^{T} |a_t - y_t|.$$

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.

1.3.1 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:

$$L(y, a) = \frac{|y - a|}{y}$$

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):

$$L(y, a) = \frac{|y - a|}{(|y| + |a|)/2}$$

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. Корень из среднеквадратичной логарифмической ошибки (root mean squared logarithmic error, RMSLE)
8. Метрики качества на основе процентных ошибок (MAPE, MdAPE, sMAPE, sMdAPE, WAPE, WMAPE, RMSPE, RMdSPE)
9. Средняя абсолютная процентная ошибка (mean absolute percentage error, MAPE)
10. Медианная абсолютная процентная ошибка (median absolute percentage error, MdAPE)
11. Симметричная средняя абсолютная процентная ошибка (symmetric mean absolute percentage error, SMAPE)
12. Симметричная медианная абсолютная процентная ошибка (symmetric median absolute percentage error, SMdAPE)
13. Взвешенная абсолютная процентная ошибка (weighted absolute percentage error, WAPE)
14. Средневзвешенная абсолютная процентная ошибка (weighted mean absolute percentage error, WMAPE)
15. Корень из среднеквадратичной процентной ошибки (root mean square percentage error, RMSPE)
16. Корень из медианной квадратичной процентной ошибки (root median square percentage error, RMdSPE)
17. Сравнение MAPE, MdAPE, sMAPE, sMdAPE, WAPE, WMAPE, RMSPE, RMdSPE
18. Метрики качества на основе относительных ошибок (MRAE, MdRAE, GMRAE)
19. Средняя относительная абсолютная ошибка (mean relative absolute error, MRAE)
20. Медианная относительная абсолютная ошибка (median relative absolute error, MdRAE)
21. Средняя геометрическая относительная абсолютная ошибка (geometric mean relative absolute error, GMRAE)

22. Сравнение MRAE, MdRAE, GMRAEV.4. Относительные метрики качества (RelMAE, RelRMSE)
23. Относительная метрика MAE (relative MAE, RelMAE)
24. Относительная метрика RMSE (relative RMSE, RelRMSE)
25. Метрики качества на основе масштабированных ошибок (MASE)
26. Средняя абсолютная масштабированная ошибка (mean absolute scaled error, MASE)

Глава 2

Свойства случайных процессов

Часть III

Свойства случайных процессов

Основано на [1], [2]

2.1 Стационарность

Определение 9.1

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 $t1, \dots, tn$, любых вещественных $x1, \dots, xn$ и любого $h > 0$:

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

Определение 9.2

Случайный процесс X_t называется стационарным 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, $t \in R$ 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. Функция $\gamma(\cdot)$ называется автоковариационной функцией (autocovariance function) и обладает следующими свойствами.

Утверждение 9.3. 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.

2.2 Эргодичность

Понятие эргодичности мотивировано законом больших чисел. Рассмотрим процесс X_t, наблюдаемый в дискретные моменты времени $t = 1, 2, \dots, T$ 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 11.1. Процесс X_t с дискретными временами $t = 1, 2, \dots$ называется эргодическим, если

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

где μ — некоторая константа.

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.

Иногда различают эргодичность по среднему, по дисперсии и т.д. 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.

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. *Панов В.* Теория случайных процессов. — 2018.