

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Локтионова Оксана Геннадьевна

Должность: проректор по учебной работе

Дата подписания: 13.09.2021 16:46:53

Уникальный программный ключ:

0b817ca911e6668abb13a5d426d79e5f1c11eabbf73e943df4a4851fda56d089

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Юго-Западный государственный университет»
(ЮЗГУ)

УТВЕРЖДАЮ

Проректор по учебной работе

О.Г. Локтионова

«10» 03

2021г.

АНАЛИЗ СИГНАЛОВ И ДАННЫХ

Методические указания по самостоятельной работе по дисциплине
«Автоматизация обработки экспериментальных данных»

Курск 2021

УДК 004.93:61

Составители: С.А. Филист

Рецензент

Доктор технических наук, профессор Р.А. Томакова

Анализ сигналов и данных: методические указания по самостоятельной работе / Юго-Зап. гос. ун-т; сост.: С.А. Филист, О.В.. Курск, 2021. 36 с.

Предназначено для студентов по дисциплине «Автоматизация обработки экспериментальных данных» по специальности 30.05.03 «Медицинская кибернетика»

Текст печатается в авторской редакции

Подписано в печать . Формат 60×84 1/16. Бумага офсетная.

Усл. печ. л. . Уч.-изд. л. . Тираж 100 экз. Заказ .

Юго-Западный государственный университет.

305040, г. Курск, ул. 50 лет Октября, 94.

Отпечатано в ЮЗГУ.

1 Вопросы и упражнения к Теме 1

На вооружении современных специалистов (ученых, инженеров, экономистов и т.д.) оказались многочисленные пакеты прикладных программ (например, Statistica, Mathcad, MatLab, Matematica и

др.), средствами которых решаются практически все возникающие в практической деятельности задачи. Однако крайне важно, чтобы пользователи этих пакетов понимали, что и как они анализируют.

В большинстве случаев обработка и анализ различных данных обусловлены необходимостью принять то или иное управленческое решение в некоторой области деятельности. Управленческое решение, как правило, опирается на анализ эмпирических сведений, почерпнутых и представленных тем или иным образом и содержащих в себе информацию, необходимую для его принятия.

Более полные теоретические сведения и ряд разобранных примеров приведены в первом модуле основного учебного пособия.

1.1 Вопросы для самопроверки

1. Почему стоит тратить усилия на изучение методов обработки и анализа экспериментальных данных, зная, что уже существует большое количество разнообразных математических программных пакетов прикладных программ?

2. Выберите одну из практических областей деятельности человека и укажите, как можно использовать методы анализа экспериментальных данных в этой области.

3. Какие бывают и что представляют собой этапы анализа данных?

4. Что такое набор данных?

5. Какими способами можно классифицировать наборы данных?

6. Почему двумерные данные представляют собой больше, чем просто два отдельных одномерных набора данных?

7. В чем разница между качественными и количественными данными?

8. Что такое переменная? Приведите несколько примеров?

9. Найдите в интернете (или журналах) табличные данные.

Определите тип данных каждой переменной.

10. Можно ли принимать управленческие решения, опираясь только на результаты, полученные в ходе анализа данных и не учитывать опыт и интуицию экспертов?

11. Какие вы знаете основные законы распределения случайных величин?

12. Как соотносятся различные законы распределения случайных величин и объекты (явления) реального мира?

13. Найдите с помощью поисковых систем в интернете несколько сайтов, посвященных различным математическим пакетам. Ознакомьтесь с презентационными материалами, представленными на этих сайтах.

1.2 Упражнения

6.1.. Найдите в сети Internet два различных набора двумерных данных. В каждом случае определите характер зависимости между двумя факторами и установите, можно ли и полезно ли иметь возможность предсказывать один фактор на основании другого.

6.2. Выполните предыдущую задачу, но для многомерных данных.

6.3. Выберите некоторую фирму и назовите две количественные переменные, представляющие интерес для нее. Укажите для каждой переменной, является она дискретной или непрерывной.

6.4. Выберите некоторую фирму и назовите две качественные переменные, представляющие для нее интерес. Для каждой переменной укажите, является она номинальной или порядковой.

6.5. Определите, вид (первичные или вторичные) следующих данных.

а) Данные правительства России о текущей экономической ситуации в каждом из субъектов федерации, используемые фирмой, планирующей расширение.

б) Данные о себестоимости продукции одного из предприятий фирмы, собранные в ходе кампании по снижению затрат».

в) Данные отчета по отрасли, приобретенные фирмой с целью оценки своего места среди конкурирующих фирм.

6.6. В таблице 1.1 содержится несколько объектов из базы данных сотрудников. Информация дана для 5 человек по состоянию на 3 июля 2008 года.

а) Что является элементарной единицей в этом наборе данных?

б) Определите вид данных: одномерные, двумерные, многомерные?

в) Какие из этих четырех переменных являются качественными, а какие – количественными?

г) Какие из переменных (если такие есть) являются порядковыми качественными переменными? Поясните свой ответ.

д) Это временной ряд или его данные об одном временном срезе?

Таблица 1.1

ол	Зарплата, руб.	Образование	Стаж, лет
	42 300	Высшее	9
	31 800	Техникум	4
	29 500	Школа	2
	58 100	Высшее	1
	36 000	Техникум	7

6.7. Рассмотрим набор данных из таблицы 1.2, содержащий информацию о некоторых видах продукции из пяти цехов (обозначаются кодами).

а) Что является элементарной единицей для этого набора

данных? б) Определите вид данных: одномерные, двумерные, многомерные?

в) Укажите качественные переменные (если они есть).

г) Есть ли в этих данных порядковая переменная? Если да, то укажите ее.

д) Это временной ряд или данные об одном временном среде?

Таблица 1.2

	Деталь	Качество	Количество служащих
-235	Тормоз	Хорошее	53
-186	Топливопровод	Отличное	37
-937	Радио	Довольно хорошее	26
-447	Шасси	Превосходное	85
-258	Провод	Хорошее	16

6.8. Что бы спланировать объем затрат на рекламу в различных средствах массовой информации (телевидение, радио, газеты и др.), вы изучаете набор данных, содержащий прошлогодние расходы каждого из ваших конкурентов на телерекламу, радио рекламу и на рекламу в газетах. Дайте полное описание типа для такого набора данных.

6.9. Объемы квартальных продаж фирмы за последние пять лет могут быть полезны для стратегического планирования.

а) Это временной ряд или данные об одном временном срезе?

б) Определите вид данных: одномерные, двумерные, многомерные?

6.10. Рассмотрим данные продаж 35 компаний.

а) Определите вид данных: одномерные, двумерные, многомерные? б) Это качественная или количественная переменная?

в) Это порядковая, номинальная или какая-либо другая

переменная?

6.11. Инспектор по контролю качества оценил каждую из произведенных сегодня партий продукции по шкале от А до Е, где А – высший сорт, а Е – низший.

а) Какал это переменная: количественная или качественная?

б) Эти порядковая, номинальная или какая-либо другая переменная?

В таблице 6.3 содержатся данные о бытовых пылесосах. а) Что является элементарной единицей в этом наборе данных? б) Это одномерные, двумерные или многомерны данные?

и) Какие из переменных являются качественными, а какие количественными?

г) Для каждой качественной переменной в этом наборе данных определите ее тип: порядковая или номинальная?

д) Это временной ряд или данные об одном временном срезе?

Таблица 1.3

	Вес, кг	Качество	Тип
170	3,5	Хорошее	Жесткий шланг
260	3,5	Отличное	Мягкий шланг
100	4,6	Хорошее	Жесткий шланг
090	3,1	Хорошее	Мягкий шланг
340	3,0	Хорошее	Мягкий шланг
120	5,2	Отличное	Жесткий шланг
130	3,5	Хорошее	Мягкий шланг

2 Вопросы и упражнения к теме 2.

Процесс распознавания включает в себя ряд этапов, одним из которых является процесс классификации. По мере расширения областей применения для систем распознавания образов расширяются области использования и алгоритмов классификации. Они встраиваются в системы диагностики, построения моделей, адаптивного и оптимального управления и др. Однако основные идеи и алгоритмы классификации лучше и полнее изложены в монографиях по математической статистике и распознаванию образов.

Более детальные теоретические сведения приведены во втором модуле учебного пособия.

2.2 Вопросы для самопроверки

1. Какова общая схема системы распознавания?
2. Можно ли обойтись без датчиков в системе распознавания образов?
3. Зачем необходимо формирование информативных признаков?
4. Почему не идут по пути создания «всесильного» классификатора? Ведь тогда можно сразу использовать исходные признаки.
5. Приведите реальные примеры распознавания образов.
6. Каков смысл «априорных вероятностей классов»?
7. Какую информацию несут условные вероятностные характеристики (при условии истинности того или иного класса) для информативных признаков?
8. Что такое апостериорные вероятности для всех рассматриваемых классов?
9. Почему мы используем решающее правило в виде максимума апостериорной вероятности?
10. Какие ещё Вам известны решающие правила?

11. Какую информацию несут условные плотности распределения (при условии истинности того или иного класса) информативных признаков?

12. Что такое средний риск?

13. Критерий минимума среднего риска соответствует ли критерию минимума вероятности ошибки классификации?

14. Запишите средний риск при дискретных информативных признаках.

15. Что такое «обучающая выборка» и зачем она необходима?

16. Что такое «самообучающиеся системы классификации»?

17. Можно ли по обучающей выборке восстановить решающую функцию напрямую?

17. Что такое «персептроны»?

18. Что общее между персептронами и нейросетями?

19.

2.2 Примеры решения задач

Пример 2.1. Имеются два класса. Заданы два ряда распределения вероятностей (при истинном первом и втором классах) и априорные вероятности классов.

Пример 2.2. Имеются два класса. Заданы два распределения вероятностей дискретных признаков. В верхней части каждой клетки таблицы По этой информации выносится решение об истинности того или иного класса. Например, если информативные признаки приняли значения: то выносится решение об истинности первого класса, ибо первая

Взвешенная вероятность 0.08 больше второй 0.06 (для этого варианта вероятность ошибки равна 0.06, а вероятность правильного решения равна 0.08). Если же информативные признаки приняли значения X решающее устройство принимает решение об истинности второго класса (для этого варианта вероятность ошибки равна 0.04, а вероятность правильного решения равна 0.12).

Суммарная вероятность

ошибки классификации равна величине

$$P(\text{ош.}) = 0.04 + 0.06 + 0.04 + 0.12 + 0.02 + 0.06 = 0.34.$$

Суммарная вероятность вынесения правильного решения равна величине $P(\text{прав. реш.}) = 0.12 + 0.08 + 0.18 + 0.14 + 0.06 + 0.08 = 0.66$.

Сумма этих вероятностей равна единице.

Очевидно, что отклонение от байесовского решающего правила (даже при одном возможном сочетании информативных признаков) приводит к увеличению вероятности вынесения ошибочного решения и одновременно к уменьшению вероятности вынесения правильного решения.

3 Вопросы и упражнения к теме 3

Целью планирования эксперимента является создание таких планов покачивания входных переменных, которые обеспечивают более быстрое и точное построение модели объекта.

Все необходимые теоретические сведения и ряд разобранных примеров приведены в третьем модуле учебного пособия.

3.1 Вопросы для самопроверки

1. Постановка задачи планирования эксперимента.
2. Построение линейной статической модели объекта при планировании эксперимента.
3. Крутое восхождение по поверхности отклика.
4. Полный факторный эксперимент.
5. Дробные реплики.
 - б. Генерирующие соотношения и определяющие контрасты для дробных реплик.
6. Обобщённый определяющий контраст для дробных реплик.
7. Насыщенные планы. Симплекс.
8. Разбиение матрицы планирования на блоки с целью устранения кусочно-постоянного дрейфа.
9. Обработка результатов эксперимента при построении линейной статической модели с использованием ортогональных планов первого порядка.
10. Ортогональное планирование второго порядка.
11. Расчёт параметров квадратичных моделей при ортогональном планировании.
12. Ротатабельное планирование.
13. Метод случайного баланса при построении матрицы планирования.

14. Выделение главных факторов с помощью диаграмм рассеяния.

Наиболее значимым фактором будет тот, у которого наибольшая помодулю дельта.

Исключим этот фактор из дальнейшего рассмотрения, введя корректировку в результаты измерения выхода. Для корректировки следует "стабилизировать" (с учетом знака). Скорректированные данные приведены в дополнительном столбце y^1 .

Повторить процедуру для оставшихся факторов.

3.2 Упражнения

8.1. Постройте ортогональный план взвешивания 7 тел и 15 тел; определите веса тел по результатам взвешивания; вычислите дисперсии получаемого веса тел (считая равноточными результаты взвешивания во всех точках плана);

8.2. Для нейтрализации кусочно-постоянного дрейфа необходимо полный факторный эксперимент 2^4 разбить на 2 блока, а затем каждый из них – тоже на 2 блока. Необходимо убедиться, что при полученном упорядоченном во времени планировании дрейф не приводит к смещению параметров модели.

8.3. Считаем, что через каждые четыре измерения аддитивный кусочно-постоянный дрейф выхода меняет свое значение. Разбивая полный факторный эксперимент 2^3 на блоки, составьте план, который не приведет к смещению параметров линейной модели за счет наличия дрейфа.

На основе матрицы планирования $2^{3 \square}$ на объекте поставлен эксперимент с одинаковым числом повторных опытов (табл. 8.У.5). Проверьте гипотезу о равной точности измерений. Если эта гипотеза принимается, то вычислите оценку дисперсии для выходной координаты. Далее постройте линейную модель объекта, оцените значимость всех параметров, незначимые параметры исключите из модели и проверьте

для полученной модели гипотезу декватности. При решении вышеуказанных задач уровень значимости

8.4. Составьте ортогональный композиционный план второго порядка при и приведите расчетные формулы для всех параметров модели, дисперсии параметров и дисперсии выхода модели.

Выделите главные факторы с использованием диаграмм рассеяния. Результаты планирования приведены в табл. 8.У.9.

8.5. По результатам планирования, приведенным в табл. 8.У.10, на основе диаграмм рассеяния выделите главные факторы.

8.6. Выделить главные факторы методом диаграмм рассеяния по результатам планирования эксперимента, приведенным в табл. 8.У.11.

8.7. Методом диаграмм рассеяния выделите главные факторы, используя результаты планирования, приведенные в табл. 8.У.12.

8.8. Выделите главные факторы с использованием диаграмм рассеяния. Результаты планирования эксперимента приведены в табл. 8.У.13.

8.9. По результатам планирования, приведенным в табл. 8.У.14, методом диаграмм рассеяния выделите главные факторы.

8.10. Методом диаграмм рассеяния выделите главные факторы, используя приведенные в табл. 8.У.15 результаты планирования эксперимента.

8.11. Выделите главные факторы объекта на основе метода диаграмм рассеяния, используя приведенные в табл. 8.У.16 результаты планирования эксперимента.

8.12. На основе метода диаграмм рассеяния выделите главные факторы, используя результаты планирования эксперимента, приведенные

8.13. Методом диаграмм рассеяния выделите главные факторы, используя результаты планирования эксперимента, приведенные в табл. 8.У.18.

4 Вопросы и упражнения к теме 4

Методы статистической обработки информации можно условно разделить на две группы: параметрические и непараметрические. Параметрические методы используют параметрические семейства зависимостей (разделяющей поверхности в распознавании образов, плотности распределения вероятности, модели объекта) и существенно используют свойства объектов. Непараметрические методы не ориентированы на указанные параметрические семейства, имеют более универсальную структуру и более широкую область применения. Они работают при большей неопределенности по априорной информации. Платой за это служит более сложная обработка исходной выборки и, как правило, непараметрические методы с этой выборкой никогда не расстаются. В лучшем случае исходная избыточная выборка заменяется укороченной выборкой, которая впитала основную информацию из исходной выборки.

Более полные теоретические сведения и ряд разобранных примеров приведены в четвертом модуле учебного пособия.

4.1 Вопросы для самопроверки

1. Чем принципиально отличаются методы параметрической и непараметрической обработки информации?
2. Есть ли в непараметрических моделях (алгоритмах) подстраиваемые параметры? Приведите примеры.
3. Как ставится задача оценивания функционалов?
4. Каковы преимущества и недостатки простейших оценок функции и плотности распределения вероятности?
5. Что такое оценка "К ближайших соседей"?
6. Что такое оценка Розенблатта – Парзена и чем она отличается от оценки "К ближайших соседей"?
7. Как добиться состоятельности оценки Розенблатта – Парзена?

6. Адаптивная перестройка оценок Розенблатта – Парзена.
9. Как на основе оценки Розенблатта – Парзена построить оценки моментов случайных величин?
10. Как на основе оценки Розенблатта – Парзена построить оценки совместной энтропии случайных величин?
11. Как на базе оценки Розенблатта – Парзена построить оценку условной плотности вероятности?
12. Адаптивный расчёт оценки условной плотности вероятности.
13. Как построить оценку прямой регрессии?
14. Можно ли на основе тех же экспериментальных данных построить оценку инверсной регрессии?
15. Как осуществляется расчёт оптимального коэффициента размытости оценки регрессии в одномерном случае.
16. Рекуррентный расчет оценки регрессии.
17. Робастные оценки регрессии.
18. Постановка задачи адаптивного управления при априорной неопределенности.
19. Основная идея построения алгоритмов адаптивного управления при априорной неопределенности.
20. Алгоритм адаптивного управления экстремальным объектом.
21. Распространение алгоритма адаптивного управления экстремальным объектом на решение задачи минимизации функций многих переменных.
22. Основные особенности применение непараметрического сглаживания при классификации в распознавании образов.

4.2 Упражнения

9.1. На основе простейшей оценки для плотности распределения вероятности вычислите оценку математического ожидания случайной величины. Рассчитайте для нее математическое ожидание и дисперсию.

9.2. На основе использования простейшей оценки для плотности распределения вероятности вычислите оценку математического ожидания от аналитической функциицентрированной случайной величины.

9.3. Используя оценку, вычисленную в предыдущем примере, найдите оценку дисперсии случайной величины.

9.4. Используя краткую выборку случайной величины и построенную на основе нее простейшую оценку плотности распределения вероятности, вычислите оценки для математического ожидания случайной величины, ее дисперсии и для математического ожидания функции случайной величины.

9.5. Используя выборку системы нескольких случайных величин, постройте простейшую оценку совместной плотности распределения вероятности и на основе нее найдите оценки элементов корреляционной матрицы.

9.6. На основе полигаммы первого порядка найдите оценку математического ожидания случайной величины. Покажите, что эта оценка является несмещенной и состоятельной.

9.7. Для оценки математического ожидания случайной величины, полученной по полигамме 1 порядка, вычислите дисперсию и сравните ее с дисперсией обычной оценки.

9.8. Придайте рекуррентный вид оценкам, рассмотренным в упражнениях 9.2, 9.3, 9.4, 9.6.

9.9. Рассчитайте оценку математического ожидания случайной величины на основе оценки Розенблатта – Парзена при треугольном виде ядра.

9.10. Рассчитайте оценку математического ожидания случайной величины на основе оценки Розенблатта – Парзена при параболическом виде ядра.

9.11. Рассчитайте оценку математического ожидания аналитической функции случайной величины на основе использования

оценки Розенблатта – Парзена при треугольном виде ядра. Используя полученный результат, получите оценку дисперсии и исследуйте ее на смещенность.

9.12. Выполните задание предыдущего примера при использовании параболического вида ядра.

9.13. На основе оценки Розенблатта – Парзена постройте оценку коэффициента корреляции.

9.14. На основе оценки Розенблатта – Парзена постройте оценку энтропии двумерной случайной величины.

9.15. Запишите оценку условной плотности распределения вероятности, используя оценки Розенблатта – Парзена для совместных плотностей вероятности (у объекта 2 входа и один выход).

9.16. Придайте рекуррентный вид оценке условной плотности распределения вероятности.

9.17. Для объекта с одним входом и одним выходом постройте непараметрические оценки прямой и обратной регрессий.

9.18. Для объекта с двумя входами и одним выходом постройте непараметрическую оценку прямой и инверсных регрессии

9.19. Для объекта с двумя входами и двумя выходами постройте непараметрические оценки прямых и инверсных регрессий.

9.20. Для объекта с одним входом и одним выходом постройте адаптивные непараметрические оценки прямой и инверсной регрессий.

9.21. Запишите алгоритм расчета робастной непараметрической оценки регрессии для объекта с двумя входами и одним выходом.

9.22. На основе оценки Розенблатта – Парзена получите оценки средней условной энтропии и среднего количества информации (объект с двумя входами и одним выходом).

9.23. Запишите алгоритм адаптивного управления

(используя непараметрические оценки инверсных регрессий) объектом с двумя управляющими входами и одним выходом.

9.24. Синтезируйте алгоритм адаптивного управления (на основе использования непараметрической оценки инверсной регрессии) для объекта с одним входом и двумя выходами.

9.25. Синтезируйте алгоритм адаптивного управления (на основе использования непараметрических оценок инверсных регрессий) объектом с двумя управляющими входами и двумя выходами.

9.26. Синтезируйте алгоритм адаптивного управления экстремальным объектом с двумя управляющими входами и двумя выходами.

9.27. Синтезируйте непараметрический алгоритм минимизации функции двух переменных.

5 Вопросы и упражнения к теме 5.

Методы дисперсионного, корреляционного и регрессионного анализов являются последовательными ступенями при исследовании связей между случайными величинами.

Методами дисперсионного анализа устанавливается наличие влияния заданного фактора на изучаемый процесс (на выходную переменную процесса) за счёт статистической обработки наблюдаемой совокупности выборочных данных. Корреляционный анализ позволяет оценить силу такой связи, а методами регрессионного анализа строится математическая модель и оценивается адекватность модели.

В последние десятилетия методы математического планирования эксперимента активно используются в методах дисперсного и регрессионного анализов и, естественно, дополняют их.

Более полные теоретические сведения и ряд разобранных примеров приведены в пятом модуле основного учебного пособия.

5.1 Вопросы для самопроверки

1. Что такое дисперсионный анализ?
2. Что такое корреляционный анализ?
3. Что такое регрессионный анализ?
4. Какие гипотезы проверяются в дисперсионном анализе?
5. Что такое статистика Фишера и критерий Фишера?
6. Основные предпосылки при решении задач спомощью дисперсионного анализа.
7. Основная идея однофакторного дисперсионного анализа.
8. Как проверяется гипотеза о равенстве нескольких дисперсий?
9. Основная идея двухфакторного дисперсионного анализа.
10. Как применяется планирование эксперимента в дисперсионном анализе?
11. Какие дисперсионные характеристики статических моделей

стохастических объектов Вам известны?

12. Что такое регрессия?

13. Как построить оценку регрессии?

14. Что такое дисперсионное отношение?

15. Приведите дисперсионные характеристики параметрических статических моделей стохастических объектов.

16. Зачем необходимо строить непараметрическую оценку регрессии?

17. Как построить оценки дисперсионных характеристик?

18. Почему некоторые модели называют субоптимальными?

5.2 Упражнения

10.1. На основе дисперсионного анализа выявить влияние фактора A на случайную величину с использованием экспериментальных данных, Представленных в таблице, при уровнях значимости: Вопросы и упражнения к модулю 6. Анализ трендов и временных рядов

Временные ряды отличаются от данных об одном временном срезе в том отношении, что в случае временных рядов сама последовательность наблюдений несет в себе важную информацию. В частности, чтобы охарактеризовать какую-либо совокупность данных в целом, вам уже недостаточно знать лишь типичное значение этих данных (например, среднее значение) или даже изменчивость этой совокупности данных (описываемую, например, стандартным отклонением). В этом случае желательно знать, что, скорее всего, произойдет дальше. Подобный прогноз должен по возможности точнее экстраполировать ближайшее поведение системы с точки зрения моделей поведения этой системы в прошлом.

Все необходимые теоретические сведения и ряд разобранных примеров приведены в шестом модуле учебного пособия.

5.3 Вопросы для самопроверки

1. В чем отличие временного ряда от данных об одном временном срезе?
2. Какая информация утрачивается, когда вы анализируете гистограмму, построенную для временного ряда?
3. Что такое "прогноз"?
4. Что такое "границы прогноза"?
5. Какую роль в прогнозировании играет математическая модель?
6. Почему анализ трендов и сезонных колебаний не позволяет получить границы прогноза?
7. Назовите четыре базовых компонента помесечных или поквартальных временных рядов (с точки зрения подхода, основанного на трендах и сезонных колебаниях).
8. Подробно опишите различия между циклическим и нерегулярным компонентами.
9. В чем отличие скользящего среднего от исходного ряда?
10. Почему в случае анализа трендов и сезонных колебаний мы используем в скользящем среднем данные именно за целый год?
11. Какие компоненты сохраняются в скользящем среднем? Какие уменьшаются или вообще исчезают?
12. Как вычисляется отношение к скользящему среднему? Какие компоненты оно представляет?
13. Что нужно сделать, чтобы на основе отношения к скользящему среднему получить сезонный индекс? Почему это возможно?
14. Что представляет собой сезонный индекс?
15. Как внести сезонную поправку в значение временного ряда? Как вы интерпретируете полученный результат?
16. Как оценивается линейный тренд в анализе трендов и сезонных колебаний?
17. Какой вид прогноза представляет линейный тренд?
18. Как получить прогноз на основе линейного тренда?

19. Какие компоненты будут представлены в этом прогнозе? Какие будут отсутствовать?
20. Каким образом гибкость ARIMA-процессов Бокса-Дженкинса помогает в анализе временных рядов?
21. Что такое "экономная модель"?
22. Как соотносится прогноз с фактическим будущим поведением оцениваемого процесса?
23. Как соотносятся границы прогноза с фактическим будущим поведением оцениваемого процесса?
24. Дайте определение процесса случайного шума с точки зрения взаимосвязи между последовательными наблюдениями.
25. Прокомментируйте следующее утверждение: если мы имеем дело с процессом случайного шума, то для его анализа не требуется применять специальные методы исследования временных рядов.
26. Что представляют собой прогноз и границы прогноза для процесса случайного шума?
27. Дайте определение процесса авторегрессии первого порядка с точки зрения взаимосвязи между последовательными наблюдениями.
28. Что представляют собой переменные X и Y в регрессионной модели для прогнозирования следующего наблюдения в процессе авторегрессии первого порядка?
29. Опишите прогнозы процесса авторегрессии в терминах последнего наблюдения и долгосрочного среднего значения для оцениваемой модели.
30. Дайте определение процесса скользящего среднего в терминах взаимосвязи между последовательными наблюдениями.
31. Какое скользящее среднее (скользящее среднее чего именно) мы имеем в виду, когда говорим о "процессе скользящего среднего"?
32. Для процесса скользящего среднего первого порядка опишите в терминах долгосрочного среднего значения для оцениваемой модели

прогнозы на два или больше периодов времени в будущее.

33. Дайте определение ARMA-процесса первого порядка в терминах взаимосвязи между последовательными наблюдениями.

34. Значение какого параметра ARMA-процесса нужно установить равным нулю, чтобы получить процесс авторегрессии?

35. Значение какого параметра ARMA-процесса нужно установить равным нулю, чтобы получить процесс скользящего среднего?

36. Опишите прогнозы на отдаленное будущее исходя из ARMA-процесса.

37. Дайте определение случайного блуждания в терминах взаимосвязи между последовательными наблюдениями.

38. Подробно опишите различия между процессом случайного шума и случайным блужданием.

39. Прокомментируйте следующее утверждение: если мы имеем дело со случайным блужданием, то для его анализа не требуется применять специальные методы исследования временных рядов.

40. Каково влияние составляющей дрейфа в случайном блуждании?

41. Опишите прогнозы для процесса случайного блуждания.

42. Чем различается поведение стационарных и нестационарных временных рядов?

43. Для каждого из перечисленных ниже видов процессов укажите, является ли он стационарным или нестационарным.

а) Процесс авторегрессии. б) Случайное блуждание.

в) Процесс скользящего среднего. г) ARMA-процесс.

44. Дайте определение ARIMA-процесса первого порядка в терминах взаимосвязи между последовательными наблюдениями.

45. Значение какого параметра ARIMA-процесса нужно установить равным нулю, чтобы получить случайное блуждание?

46. Как получить ARMA-процесс из ARIMA-процесса?

47. Опишите прогнозы на отдаленное будущее исходя из

ARIMA-процесса.

5.4 Упражнения

11.1. Для каждого из перечисленных ниже случаев укажите, присутствует ли в нем значительный сезонный компонент. Поясните свой ответ.

а) Продажа цветной оберточной бумаги (объемы продаж фиксируются ежемесячно).

б) Количество авиапассажиров, направляющихся из Красноярска в Сочи (количество пассажиров фиксируется ежемесячно).

в) Биржевой индекс (фиксируется ежедневно). Предполагается, что биржа работает эффективно, в результате чего любые прогнозируемые тенденции уже устранены действиями крупных инвесторов, пытающихся извлечь из них для себя выгоду.

11.2. Некоторое время вас терзают подозрения, что проблемы с производством обостряются, как правило, именно в зимние месяцы в первом квартале каждого года. Анализ трендов и сезонных колебаний процента производственного брака позволил установить следующие значения сезонных индексов: 1,00 – 1-й квартал; 1,01 – 2-й квартал; 1,03 – 3-й квартал и 0,97 – 4-й квартал. Подтверждает ли этот анализ ваши подозрения о том, что наивысший процент производственного брака приходится именно на первый квартал? Если да, обоснуйте свой ответ. Если нет, тогда может быть, следует обратить внимание на какой-то другой квартал?

11.3. В январе у одного из банков зафиксировано 38 091 операции в сети автоматических кассовых аппаратов, а в феврале 43 182. Соответствующий сезонный индекс для января равен 0.925, а для февраля — 0.986.

а) На какой процент увеличилось количество операций в сети

автоматических кассовых аппаратов с января по февраль?

б) На какой процент должно было бы, по вашему мнению, увеличиться количество операций в сети автоматических кассовых аппаратов с января по февраль? (Подсказка: воспользуйтесь сезонными индексами.)

в) Определите, учитывал сезонную поправку, количество операции в сети автоматических кассовых аппаратов для каждого из этих двух месяцев.

г) На какой процент увеличилось (или уменьшилось) количество операций в сети автоматических кассовых аппаратов с января по февраль с учетом сезонной поправки?

11.4. На производственном собрании все выразили удовлетворение тем фактом, что объем продаж в фирме вырос с 21 791 000 рублей в третьем квартале до 22 675 000 рублей в четвертом квартале. Кратко опишите анализ этой ситуации (с учетом поправки па сезон), если вам известно, что сезонный индекс для третьего квартала равен 1,061, а для четвертого – 1,180. Так ли радужна картина с объемами продаж в фирме, как показалось участникам собрания?

11.5. В таблице 11.1 представлены поквартальные величины нетто-продажи (суммарные продажи компании за вычетом возврата продукции, штрафов, расходов по доставке, скидок и т.п.) и доходы компании крупного производителя сельскохозяйственного и промышленного оборудования.

а) Постройте график временного ряда для этой совокупности данных. Опишите все тенденции и сезонные колебания, замеченные вами на этом графике.

б) Вычислите скользящее среднее (используя каждый раз данные за один год) для этого временного ряда. Постройте график временного ряда, содержащий и данные, и скользящее среднее.

в) Найдите сезонный индекс для каждого квартала. Кажутся ли полученные вами значения обоснованными, если исходить из построенного графика временного ряда?

г) Какой из кварталов (1, 2, 3 или 4) оказывается для компании самым неблагоприятным? Насколько ниже (в среднем) оказывается объем продажи в это квартале по сравнению с типичным кварталом в течение года?

д) Определите значения объемов продаж с поправкой на сезон, соответствующие каждому из исходных величин объемов продаж.

е) С третьего по четвертый квартал 2005 г. объемы продаж увеличились с 2673 до 2718. Как выглядит картина с учетом сезонной поправки?

ж) Со второго по третий квартал 2007 г. объемы продаж компании снизились с 3521 до 3430. Как выглядит картина с учетом сезонной поправки? з) Найдите уравнение регрессии для прогнозирования долгосрочного тренда изменения объемов продаж (с учетом сезонной поправки) для каждого периода времени, используя в качестве значений переменной X числа 1, 2,

11.6. В таблице 11.2 приведены данные о квартальных объемах продажи международной компании, специализирующейся на производстве известных марок продуктов питания. В годовом отчете за 2003 г. утверждается, что "значительное влияние на ежеквартальные результаты деятельности компании оказывают сезонные факторы, неразрывно связанные с ее бизнесом".

а) Постройте график временного ряда для этой совокупности данных. Согласны ли вы, что в этом случае действительно имеют место сезонные факторы?

б) Вычислите скользящее среднее (используя каждый раз данные за один год) для этого временного ряда. Постройте график временного ряда, включающий как данные, так и значения скользящего среднего.

в) Опишите циклическое поведение (если оно наблюдается) скользящего среднего.

г) Найдите сезонный индекс для каждого квартала. Кажутся ли полученные вами значения обоснованными, если исходить из построенного

графика временного ряда?

д) Какой из кварталов (1, 2, 3 или 4) оказывается для компании самым благоприятным? Насколько в среднем выше объем продажи в этом квартале по сравнению с типичным кварталом в течение года?

е) Какой из кварталов (1, 2, 3 или 4) оказывается для компании самым неблагоприятным? Насколько в среднем ниже оказывается объем продажи в этом квартале по сравнению с типичным кварталом в течение года?

ж) Определите значения объемов продажи с учетом сезонной поправки соответствующие каждой из исходных величин объема продажи. Постройте график для этого временного ряда с поправкой на сезон.

з) Опишите поведение этого временного ряда с поправкой на сезон. В частности, выявите любые: изменения непостоянства продаж за этот период времени. Исходя из накопленных за несколько прошлых лет данных были обнаружены сезонные колебания объемов продажи в некоторой фирме. Сезонный индекс за ноябрь равен 1.08; за декабрь – 1.38 и за январь — 0.84. Объем продажи в ноябре составил 285 167 рублей.

а) Можно ли, как правило, ожидать увеличения объемов продажи с ноября по декабрь в «типичном» году. Обоснуйте свой ответ.

б) Найдите объем продажи в ноябре с поправкой на сезон.

в) Внесите в показатель объема продажи в ноябре (с поправкой на сезон) сезонность с помощью декабрьского индекса, чтобы найти ожидаемый объем продажи в декабре-

г) Было объявлено, что объем продажи в декабре составил 430 106 млн. руб. Оказался, ли этот показатель выше или ниже, чем ожидалось, если исходить из объема продажи в ноябре?

д) Найдите объем продажи в декабре с поправкой на сезон.

е) Объемы продажи с ноября по декабрь – с учетом поправки на сезон – выросли или, наоборот, снизились? О чем это свидетельствует?

ж) Пользуясь тем же методом, что и в п. "в", найдите ожидаемый объем продаж в январе исходя из объема продажи в декабре.

11.7. Вы решили изучить поквартальное количество посетителей ресторана для любителей горнолыжного спорта, воспользовавшись методом анализа трендов и сезонных колебаний. Квартальные сезонные индексы равны 1,45; 0,55; 0,72 и 1,26 для 1-го, 2-го, 3-го и 4-го кварталов соответственно. Линейный тренд оценивается уравнением вида $5\,423 + 408(\text{номер квартала})$, причем номер квартала начинается с 1 в первом квартале 2001 г. и увеличивается на единицу для каждого последующего квартала.

а) Найдите прогнозируемое значение (с поправкой на сезон) для первого квартала 2005 г.

б) Найдите прогнозируемое значение (с поправкой на сезон) для второго квартала 2005 г.

в) Почему прогнозируемое значение с поправкой на сезон оказалось большим во втором квартале, в котором, как можно было бы предположить, ресторан посещает меньшее количество лыжников?

г) Найдите прогнозируемое значение для первого квартала 2006 г. д) Найдите прогнозируемое значение для второго квартала 2006 г.

е) С учетом поправки на сезон и в соответствии с оценкой линейного тренда ответьте на вопрос: насколько больше посетителей ожидается обслуживать в ресторане каждый квартал в сравнении с предыдущим кварталом?

ж) Стратегический бизнес-план включает проект значительное расширения ресторанного бизнеса (количество посетителей ресторана должно достичь 70 000 за год). В каком году – в соответствии с прогнозом это должно произойти впервые? (Подсказка: вычислите и сложите четыре прогнозируемых значения для каждого года, чтобы найти годовые итоговые показатели для 2006 и 2006 гг.).

11.8. Какой тип анализа временных рядов обеспечит получение простейших результатов для изучения спроса на мазут (используемый для обогрева), который, как правило, достигает пика в зимний

период?

11.9. Для каждой из перечисленных ниже ситуаций укажите, какому типу процесса (стационарный или нестационарный) она соответствует.

а) Цена, одной акции компании IBM, фиксируемая ежедневно.

б) Прайм-рейт, фиксируемый еженедельно и представляющий собой публикуемую байками процентную старку по кредитам для наилучших заемщиков.

в) Толщина бумаги измеряемая пять раз в минуту в процессе производства бумаги и ее намотки на рулоны. (Предполагается, что этот процесс находится под контролем.)

г) Цена одной страницы рекламного объявления в журнале TV Guide; изменяется раз в год.

11.10. Выберите какую-либо интересующую вас фирму и получите данные о поквартальных объемах продаж этой фирмы по крайней мере за три последовательных года (для этого можно воспользоваться ежегодными отчетами фирмы, которые можно получить в библиотеке или через Internet).

а) Изобразите график временного ряда и прокомментируйте структуру, которая следует из этого графика.

б) Вычислите скользящее среднее за год, отобразите его на своем графике и прокомментируйте,

в) Вычислите сезонные индексы, отобразите их на своем графике и прокомментируйте.

г) Вычислите и отобразите на своем графике временной ряд с поправкой на сезон, затем прокомментируйте полученный результат. В частности, ответьте на вопрос: какую новую информацию можно извлечь в результате внесения сезонной поправки?

д) Вычислите линию тренда и внесите в нее сезонную поправку, чтобы получить прогнозы на два последующих. Отобразите эти

прогнозы на своем графике наряду с исходными данными.
Прокомментируйте, насколько правдоподобными кажутся вам эти прогнозы.

6 Вопросы и упражнения к теме 6.

Идентификация – это процесс построения моделей объектов различной природы. Теория идентификации имеет в своем арсенале достаточно эффективные методы и алгоритмы, на базе которых разработаны и широко используются программные комплексы.

Процесс идентификации складывается из двух взаимосвязанных этапов: идентификации структуры моделей и идентификации параметров в моделях выбранной структуры. При построении структуры модели (или набора конкурирующих либо взаимодополняющих структур) используется априорная информация об объекте. Для каждого класса объектов формируются банки структур с сопутствующей информацией.

Более полные теоретические сведения и ряд разобранных примеров приведены в седьмом модуле основного учебного пособия.

6.1 Вопросы для самопроверки

1. Что такое идентификация?
2. Постановка задачи подстройки параметров нелинейных моделей.
3. Критерий наименьших квадратов.
4. Метод наименьших квадратов при линейной параметризации модели.
5. Как рассчитывается корреляционная матрица для вектора параметров модели?
6. Как построить доверительный интервал для параметров модели?
7. Как вычислить дисперсию выхода модели
8. Как проверить гипотезу адекватности модели (при некоррелированных равноточных измерениях выхода)?
9. Постройте линейную модель для объекта с одним входом

и одним выходом и исследуйте её свойства.

10. Постройте линейную модель для объекта с двумя входами и одним выходом.

11. Опишите метод последовательной линеаризации при подстройке параметров статических моделей.

12. Метод последовательной линеаризации при подстройке параметров на основе критерия наименьших квадратов.

13. Как вычислять робастные оценки параметров?

14. Запишите критерий, которому удовлетворяет оценка медианы?

15. Адаптивные алгоритмы метода наименьших квадратов при линейной параметризации модели и некоррелированных измерениях выхода объекта.

16. Адаптивные алгоритмы метода наименьших квадратов при линейной параметризации модели и забывании информации.

17. Адаптивные алгоритмы метода наименьших квадратов при нелинейной параметризации модели и некоррелированных измерениях выхода объекта.

18. Адаптивные алгоритмы при подстройке нестационарных параметров моделей.

19. Адаптивные алгоритмы подстройки робастных оценок параметров моделей.

7 Вопросы и упражнения к теме 7.

В динамическом режиме поведение объектов описывается различными динамическими уравнениями: обыкновенными дифференциальными, интегральными, интегродифференциальными уравнениями; уравнениями с запаздываниями; уравнениями в частных производных и их дискретными аналогами. С целью упрощения изложения материала будут рассматриваться наиболее простые дискретные модели. Последние выбраны именно потому, что получаемые алгоритмы идентификации и управления напрямую реализуемы на цифровой вычислительной технике (мини-, микро-ЭВМ, микропроцессоры).

В адаптивных системах обработки информации и управления происходит приспособление к изменяющимся условиям и неизвестным характеристикам объекта.

Задается (или синтезируется) структура модели с точностью до параметров. На базе нее из критериев оптимальности синтезируется управление, которое, естественно, зависит от параметров модели. Параметры модели перестраиваются блоком идентификации непрерывно по мере поступления новой информации об объекте. Более полные теоретические сведения и ряд разобранных примеров приведены в восьмом модуле учебного пособия.

7.1 Вопросы для самопроверки

1. Почему предпочитают брать за основу дискретные динамические модели по сравнению с непрерывными?
2. Какой применяют критерий при получении оптимальной структуры модели для линейных стохастических объектов?
3. Какова структура оптимальной модели линейных стохастических объектов?
4. Что такое «окрашенная помеха»?
5. Что такое «дискретный белый шум»?

6. Как конструируют субоптимальные структуры моделей нелинейных стохастических объектов?

7. Как по оптимальной (или субоптимальной) структуре модели можно создать структуру модели с переменными перестраиваемыми (по мере поступления новой информации о входах и выходах объекта) параметрами?

8. На примере продемонстрируйте принцип построения итеративной модели с переменными перестраиваемыми параметрами.

9. Какова идея построения модели с перестраиваемыми параметрами при применении функций чувствительности?

10. Что такое функции чувствительности?

11. Объясните принцип построения уравнений чувствительности.

12. На некотором простом примере продемонстрируйте применение простейшего адаптивного алгоритма для перестройки параметров модели с функциями чувствительности.

13. Как оцениваются параметры моделей, если неизвестные параметры стохастических объектов являются нестационарными?

14. На некотором примере продемонстрируйте применение простейшего адаптивного алгоритма для перестройки параметров итеративной модели.

15. Поставьте задачу адаптивного управления стохастическими динамическими объектами.

16. Что такое адаптивные системы с идентификатором?

17. Какова основная идея построения алгоритмов адаптивного управления с идентификатором?

18. Что такое локальный критерий оптимальности?

19. Приведите примеры синтеза устройств управления для простейших линейных систем.

20. Причины дополнительного прогнозирования параметров

моделей при синтезе алгоритмов адаптивного управления с идентификатором?

21. Какова общая схема синтеза алгоритмов адаптивного управления для обычных линейных систем?

22. Как синтезируются алгоритмы адаптивного управления для обычных нелинейных систем?

23. Чем отличается процесс синтеза алгоритмов адаптивного управления для динамических систем с чистыми запаздываниями?

24. Приведите пример синтеза устройства управления для простейшей линейной динамической системы с чистым запаздыванием.

необходимо построить оптимальную структуру модели и вычислить параметры модели с использованием рекурсивной модели на основе:

- а) рекуррентного алгоритма наименьших квадратов; в) простейшего адаптивного алгоритма;
- д) алгоритма Поляка.

Библиографический список

1. Рубан, А. И. Методы анализа данных. Учебное пособие / А. И. Рубан Изд. 2-е., исправл. и доп. Красноярск: ИПЦ КГТУ, 2004. – 319 с.
2. Кнут, Д. Э. Искусство программирования том 2. Получисленные алгоритмы / Д. Э. Кнут, Изд 3-е., Пер. с англ. : Уч. пособие. – М.: Издательский дом «Вильямс», 2000. – 832 с.
3. Красовский, Г. И. Планирование эксперимента / Г. И. Красовский, Г. Ф. Филаретов, Минск: Изд-во БГУ, 1982. – 302 с.
4. Вентцель, Е. С. Теория вероятностей / Е. С. Вентцель. – М.: Высш. шк., 2001. – 575 с.
5. Рубан, А. И. Теория вероятностей и математическая статистика / А. И. Рубан. Красноярск: ИПЦ КГТУ, 2002. – 320 с.
6. Гмурман, В. Е. Руководство к решению задач по теории вероятностей и математической статистике: Учеб. пособие для студентов вузов / В. Е. Гмурман. Изд. 5-е, стер. М.: Высш. шк., 1999. – 400 с.
7. Гмурман, В. Е. Теория вероятностей и математическая статистика: Учеб. пособие для студентов вузов / В. Е. Гмурман. М.: Высш. шк., 1977. – 479 с.
8. Сигел, Э. Ф. Практическая бизнес-статистика / Э. Ф. Сигел. М.: Издательский дом «Вильямс», 2002. – 1056 с.
9. Боровиков, В. П. STATISTICA. Искусство анализа данных на компьютере: Для профессионалов. / В. П. Боровиков. Спб.: Питер, 2003. – 688с.6.3.
10. Боровиков, В. П. Популярное Введение в программу Statistica / В. П. Боровиков. Спб.: Питер, 2002. – 301 с.
11. Бендат, Дж. Прикладной анализ случайных данных / Дж. Бендат, А. Пирсол. М.: Мир, 1989. – 540 с.