

فهرست مطالب

۱۲..... مقدمه ناشر

۱۳..... پیش‌گفتاری بر «ملزومات آماری در علم داده‌ها».....

فصل صفر

۱۵..... کتاب‌شناسی و تقدیم به

۱۶..... قواعد مورد استفاده در این کتاب

۱۷..... اصطلاحات کلیدی و نمادهای استفاده‌شده در کتاب

۱۷..... استفاده از مثال‌های کد.....

۱۸..... تقدیم به

فصل اول

۱۹..... تحلیل اکتشافی داده‌ها.....

۲۰..... عناصر داده‌های ساخت‌یافته

۲۲..... مطالعه بیش‌تر

۲۳..... داده‌های مستطیلی

۲۴..... دیتافریم‌ها و اندیس‌ها

۲۵..... ساختارهای داده‌های غیرمستطیلی.....

۲۶..... برآورد مکان

۲۷..... خلاصه‌سازی داده‌ها.....

۲۷..... میانگین

۲۸..... میانه و برآوردهای مقاوم

۲۹..... داده‌های پرت

۳۰..... مثال: برآوردهای مکان جمعیت و نرخ جرم و جنایت

۳۱..... مطالعه بیش‌تر

۳۱..... برآورد پراکندگی.....

۳۳..... انحراف استاندارد و برآوردهای مربوط به آن

۳۵..... برآوردهای مبتنی بر صدک

۳۶..... مثال: برآوردهای پراکندگی جمعیت ایالتی

۳۷..... مطالعه بیش‌تر

۳۷.....	کشف توزیع داده
۳۸.....	صدها و نمودارهای جعبه‌ای
۴۰.....	جداول فراوانی و هیستوگرام‌ها
۴۲.....	نمودارهای چگالی و برآوردها
۴۴.....	مطالعه بیش تر
۴۵.....	کشف داده‌های دودویی و دسته‌ای
۴۷.....	داده‌های «عددی» به‌عنوان داده‌های «دسته‌ای»
۴۷.....	مد
۴۷.....	امید ریاضی
۴۸.....	احتمال
۴۸.....	همبستگی
۵۲.....	نمودارهای نقطه‌ای
۵۳.....	مطالعه بیش تر
۵۳.....	اکتشاف دو یا چند متغیر
۵۴.....	انبارک شش ضلعی و کانتورها (رسم داده‌های عددی در مقابل داده‌های عددی)
۵۶.....	دو متغیر دسته‌ای
۵۸.....	داده‌های دسته‌ای و عددی
۶۰.....	نمایش چند متغیر
۶۲.....	مطالعه بیش تر
۶۲.....	خلاصه

فصل دوم

۶۳.....	توزیع داده و نمونه‌گیری
۶۵.....	نمونه‌گیری تصادفی و بایاس نمونه
۶۶.....	بایاس نمونه‌گیری خودگزین
۶۷.....	بایاس
۶۸.....	انتخاب تصادفی
۶۸.....	اندازه در مقابل کیفیت: چه زمانی اندازه مهم است؟
۶۹.....	میانگین نمونه در مقابل میانگین جمعیت
۷۰.....	مطالعه بیش تر
۷۰.....	بایاس انتخاب
۷۲.....	رگرسیون به میانگین
۷۳.....	مطالعه بیش تر
۷۳.....	توزیع نمونه‌گیری یک آماره
۷۶.....	نظریه حد مرکزی

۷۶	خطای استاندارد
۷۷	مطالعه بیش تر
۷۸	بوت استرپ
۸۱	نمونه‌گیری مجدد در مقابل بوت استرپ
۸۱	مطالعه بیش تر
۸۲	بازه‌های اطمینان
۸۴	توزیع نرمال
۸۵	نمودارهای QQ و نرمال استاندارد
۸۷	توزیع‌های دم‌طولانی
۸۹	مطالعه بیش تر
۸۹	توزیع t دانش‌آموز
۹۱	مطالعه بیش تر
۹۱	توزیع دوجمله‌ای
۹۴	مطالعه بیش تر
۹۴	توزیع مربع-کای
۹۵	مطالعه بیش تر
۹۵	توزیع F
۹۵	مطالعه بیش تر
۹۶	پواسون و توزیع‌های مربوطه
۹۶	توزیع پواسون
۹۷	توزیع نمایی
۹۷	تخمین نرخ شکست (خرابی)
۹۸	توزیع ویبول
۹۹	مطالعه بیش تر
۹۹	خلاصه

فصل سوم

۱۰۰	آزمایش‌های علمی و آزمون معناداری
۱۰۱	آزمون A/B
۱۰۴	چرا باید گروه کنترل داشته باشیم؟
۱۰۵	چرا فقط A/B؟ چرا C/D، ... نه؟
۱۰۶	مطالعه بیش تر
۱۰۶	آزمون‌های فرضیه
۱۰۷	فرض صفر
۱۰۸	فرضیه جایگزین

۱۰۸	آزمون‌های یک‌طرفه در مقابل آزمون‌های دوطرفه
۱۰۹	مطالعه بیشتر
۱۰۹	نمونه‌گیری مجدد
۱۱۰	آزمون جایگشت
۱۱۱	مثال: جذابیت وب
۱۱۵	آزمون‌های جایگشت بوت استرپ و کامل
۱۱۵	آزمون‌های جایگشت: خط پایان علوم داده
۱۱۶	مطالعه بیشتر
۱۱۶	معناداری آماری و مقادیر p
۱۱۹	مقدار p
۱۲۰	آلفا
۱۲۱	جدال بر سر مقدار p
۱۲۲	معناداری کاربردی
۱۲۲	خطاهای نوع ۱ و نوع ۲
۱۲۲	علوم داده و مقدار p
۱۲۳	مطالعه بیشتر
۱۲۳	آزمون‌های t
۱۲۵	مطالعه بیشتر
۱۲۵	آزمون چندگانه
۱۲۹	مطالعه بیشتر
۱۲۹	درجه آزادی
۱۳۰	ANOVA
۱۳۴	آماره F
۱۳۵	ANOVA دوطرفه
۱۳۶	مطالعه بیشتر
۱۳۶	آزمون مربع کای
۱۳۶	آزمون مربع کای: یک روش نمونه‌گیری مجدد
۱۳۹	آزمون مربع کای: نظریه آماری
۱۴۰	آزمون دقیق فیشر
۱۴۲	ارتباط برای علوم داده
۱۴۳	مطالعه بیشتر
۱۴۳	الگوریتم راهزن چنددست
۱۴۶	مطالعه بیشتر
۱۴۶	قدرت و اندازه نمونه
۱۴۸	اندازه نمونه

مطالعه بیش تر ۱۵۰

خلاصه ۱۵۰

فصل چهارم

رگرسیون و پیش بینی ۱۵۱

رگرسیون خطی ساده ۱۵۲

معادله رگرسیون ۱۵۳

مقادیر برازش شده و مانده ها ۱۵۶

حداقل مربعات ۱۵۸

پیش بینی در مقابل توضیح (پروفایلینگ) ۱۵۹

رگرسیون خطی چندگانه ۱۵۹

مثال: داده های خانه های کینگ کانتی ۱۶۰

ارزیابی مدل ۱۶۲

اعتبارسنجی متقابل ۱۶۶

انتخاب مدل و رگرسیون گام به گام ۱۶۷

رگرسیون وزن دار ۱۷۱

مطالعه بیش تر ۱۷۳

پیش بینی با استفاده از رگرسیون ۱۷۳

خطرات برون یابی ۱۷۳

بازه های اطمینان و پیش بینی ۱۷۴

متغیرهای عامل در رگرسیون ۱۷۵

نمایش متغیرهای ساختگی ۱۷۶

متغیرهای عامل با چندین سطح ۱۷۹

متغیرهای عاملی مرتب ۱۸۱

تفسیر معادله رگرسیون ۱۸۲

پیشگوهای همبسته ۱۸۲

هم خطی چندگانه ۱۸۴

متغیرهای مختلط ۱۸۴

تعامل ها و اثرات مهم ۱۸۶

تشخیص های رگرسیون ۱۸۸

داده های پرت ۱۸۹

مقادیر تأثیر گذار ۱۹۱

ناهم واریداری، نرمال نبودن و خطاهای همبسته ۱۹۳

نمودارهای مانده جزئی و غیرخطی بودن ۱۹۶

رگرسیون اسپایلین و چند جمله ای ۱۹۸

۱۹۹	چند جمله‌ای
۲۰۱	اسپیلاین
۲۰۲	مدل‌های جمعی تعمیم یافته
۲۰۴	مطالعه بیش تر
۲۰۴	خلاصه

فصل پنجم

طبقه‌بندی ۲۰۵

۲۰۷	نایو بیز
۲۰۸	چرا طبقه‌بندی بیزین دقیق غیر کاربردی است؟
۲۰۸	راه حل نایو
۲۱۱	متغیرهای پیشگوی عددی
۲۱۲	مطالعه بیش تر
۲۱۲	تحلیل افتراقی
۲۱۳	ماتریس کوواریانس
۲۱۴	تفکیک کننده خطی فیشر
۲۱۴	مثال
۲۱۷	مطالعه بیش تر
۲۱۸	رگرسیون لجستیک
۲۱۸	تابع پاسخ لجستیک و لاجیت
۲۲۰	رگرسیون لجستیک و GLM
۲۲۱	مدل‌های خطی تعمیم یافته
۲۲۲	مقادیر پیش‌بینی شده رگرسیون لجستیک
۲۲۳	تفسیر ضرایب و نسبت شانس
۲۲۴	رگرسیون خطی و لجستیک: شباهت‌ها و تفاوت‌ها
۲۲۴	برازش مدل
۲۲۵	ارزیابی مدل
۲۲۷	تحلیل مانده‌ها
۲۲۹	مطالعه بیش تر
۲۲۹	ارزیابی مدل‌های طبقه‌بندی
۲۳۰	ماتریس اغتشاش
۲۳۲	مسأله کلاس نادر
۲۳۳	صحت، یادآوری و ویژگی
۲۳۴	منحنی ROC
۲۳۶	AUC

۲۳۷	لیفت
۲۳۹	مطالعه بیش تر
۲۳۹	استراتژی‌هایی برای داده‌های نامتوازن
۲۴۰	نمونه برداری کاهش
۲۴۲	نمونه برداری افزایشی یا وزن دهی افزایشی/کاهش
۲۴۳	تولید داده
۲۴۴	طبقه بندی مبتنی بر هزینه
۲۴۴	بررسی پیش بینی ها
۲۴۶	خلاصه

فصل ششم

۲۴۷	یادگیری ماشین آماری
۲۴۹	k- نزدیک ترین همسایه
۲۵۰	مثال: پیش بینی نکول
۲۵۲	معیارهای فاصله
۲۵۳	رمز گذار وان هات
۲۵۴	استانداردسازی (نرمال سازی، امتیاز z)
۲۵۷	انتخاب K
۲۵۸	KNN به عنوان موتور ویژگی
۲۶۰	مدل های درختی
۲۶۲	مثال
۲۶۴	الگوریتم پارتیشن بندی بازگشتی
۲۶۶	اندازه گیری همگن بودن یا خلوص
۲۶۸	جلوگیری از رشد درخت
۲۶۸	کنترل پیچیدگی در R
۲۶۹	کنترل پیچیدگی درخت در پایتون
۲۶۹	پیش بینی یک مقدار پیوسته
۲۷۰	درخت ها چگونه استفاده می شوند؟
۲۷۱	بگینگ و جنگل تصادفی
۲۷۲	بگینگ
۲۷۳	جنگل تصادفی
۲۷۷	اهمیت متغیر
۲۸۰	فراپارامترها
۲۸۱	بوستینگ
۲۸۲	الگوریتم بوستینگ

۲۸۳XGBOOST
۲۸۶ lightgbm
۲۸۸Catboost
۲۸۹اجتناب از بیش‌برازش
۲۹۴فراپارامترها و اعتبارسنجی متقابل
۲۹۷خلاصه

فصل هفتم

۲۹۸..... یادگیری بدون نظارت

۳۰۰تحلیل مؤلفه اساسی
۳۰۱ مثال
۳۰۴محاسبه مؤلفه‌های اساسی
۳۰۷تحلیل تناظر
۳۰۹خوشه‌بندی
۳۰۹k-means خوشه‌بندی
۳۱۰ مثال
۳۱۳الگوریتم k-means
۳۱۴تفسیر خوشه‌ها
۳۱۵انتخاب تعداد خوشه‌ها
۳۱۷ارزیابی خوشه‌بندی
۳۱۸الگوریتم Kmeans++
۳۲۰الگوریتم Mini-batch kmeans
۳۲۱خوشه‌بندی DBSCAN
۳۳۰خوشه‌بندی سلسله‌مراتبی
۳۳۱ مثال
۳۳۲دندوگرام
۳۳۴الگوریتم تجمعی
۳۳۴معیارهای عدم تشابه
۳۳۶خوشه‌بندی مبتنی بر مدل
۳۳۶توزیع نرمال چندمتغیره
۳۳۸ترکیب نرمال‌ها
۳۴۰انتخاب تعداد خوشه‌ها
۳۴۳مقیاس‌بندی و متغیرهای دسته‌ای
۳۴۳روش‌های مختلف برای مقیاس‌بندی
۳۴۴تأثیر مقیاس‌بندی خوشه‌بندی