

## فهرست مطالب

۷	مقدمه ناشر
۱۰	<b>فصل اول</b> تجزیه و تحلیل داده‌ها
۱۰	عناصر داده‌های ساخت‌یافته
۱۱	داده‌های مستطیلی
۱۳	چارچوب‌ها و شاخص‌های داده
۱۳	ساختارهای داده غیر مستطیلی
۱۵	خلاصه در علم داده
۱۶	تخمین موقعیت مکانی
۱۷	میانگین
۱۷	میانگین پیراسته یا میانگین اصلاحی
۱۸	میانگین وزنی
۱۸	میانه (Median)
۲۰	میانه وزنی (Weighted Mean)
۲۱	مقادیر پرت (Outliers)
۲۱	تخمین تنوع و پراکندگی
۲۱	واریانس $\sigma^2$
۲۲	انحراف معیار (STD)
۲۲	تفاوت واریانس و انحراف معیار
۲۳	میانگین قدرمطلق انحراف (Mean absolute Deviation)
۲۴	میانه قدرمطلق انحراف (Median Absolute Deviation - MAD)
۲۴	تشخیص ناهنجاری (Anomaly Detection)
۲۵	ناهنجاری نقطه‌ای (Point Anomaly)
۲۵	ناهنجاری زمینه‌ای (Contextual Anomaly)
۲۶	روش‌های بصری ساده برای تشخیص ناهنجاری
۲۸	روش‌های تشخیص ناهنجاری با استفاده مدل‌ها و توابع
۳۱	Range
۳۲	چندک در علم داده (Quantile)
۳۳	IQR (Interquartile Range) یا فاصله بین چارکی
۳۶	نما (Mode)
۳۶	گشتاور در آمار
۳۷	نمودارهای میله‌ای
۳۸	نمودارهای دایره‌ای

۳۹	نمودارهای خطی
۴۰	نمودارهای پراکنندگی
۴۱	نمودارهای جعبه‌ای
۴۲	نمودارهای حرارتی
۴۳	نمودار هیستوگرام (Histogram) در پایتون
۴۴	نمودار هگزین Hexbin Plot
۴۶	همبستگی (Correlation)
۴۷	همبستگی خطی
۴۹	روش تشخیص نوع همبستگی
۵۱	احتمال در آمار
۵۱	مفاهیم کلیدی در احتمال
۵۲	قوانین اساسی احتمال
۵۳	قانون جمع احتمال برای رویدادهای مستقل
۵۴	قانون جمع احتمال برای رویدادهای غیرمستقل
۵۴	امید ریاضی
۵۵	خلاصه

## فصل دوم داده‌ها و توزیع‌های نمونه‌گیری ۵۶

۵۶	نمونه‌گیری تصادفی و سوگیری نمونه
۵۶	سوگیری (Bias)
۵۸	انتخاب تصادفی
۵۹	اندازه در مقابل کیفیت
۶۰	چه زمانی به مقادیر انبوه داده نیاز است؟
۶۰	آشنایی با توزیع نمونه‌ای
۶۱	تفاوت توزیع نمونه با توزیع جمعیت
۶۲	خطای استاندارد (Standard Error)
۶۲	تفاوت خطای استاندارد و انحراف استاندارد
۶۳	محاسبه خطای استاندارد با یک نمونه یا چند نمونه
۶۵	بوت استرپ (Bootstrapping) چیست؟
۶۸	فاصله اطمینان (Confidence Interval - CI)
۷۲	نمودار چندک-چندک Q-Q-PLOT
۷۵	درجه آزادی (df) چیست؟
۷۷	چولگی (Skewness)
۷۷	کشیدگی (Kurtosis)
۷۸	توزیع نرمال
۸۴	توزیع یونیفرم (یکنواخت)

۸۶	توزیع برنولی
۸۸	توزیع چند جمله‌ای (مالتی نومینال)
۹۰	توزیع آماری نوع t - (t-Distribution)
۹۴	توزیع دو جمله‌ای
۹۵	تابع جرم احتمال PMF
۹۶	تابع توزیع تجمعی CDF
۹۷	تابع چگالی احتمال (PDF)
۹۹	توزیع نمایی (Exponential Distribution)
۱۰۴	توزیع پواسون
۱۰۷	توزیع خی دو (Chi-Square)

### فصل سوم آزمایش‌های آماری و آزمون فرض آماری

۱۰۹	آزمون آماری و آزمون فرض آماری
۱۱۱	آزمون‌های فرضیه
۱۱۲	فرضیه صفر
۱۱۳	نمونه‌گیری مجدد
۱۱۳	تست A/B
۱۱۷	تست جایگشت - جابجایی (Permutation Test)
۱۲۴	آزمون t
۱۲۸	آزمون خی دو (Chi-Square)
۱۳۲	آزمون فیشر - F
۱۳۷	آزمون پواسون
۱۴۱	یادگیری ماشین کلاسیک چیست؟
۱۴۱	انواع یادگیری ماشین
۱۴۱	۱. یادگیری تحت نظارت (Supervised Learning)
۱۴۲	۲. یادگیری بدون نظارت (Unsupervised Learning)
۱۴۳	۳. یادگیری تقویتی (Reinforcement Learning)
۱۴۴	آینده یادگیری ماشین
۱۴۶	زمینه‌های نوظهور
۱۴۶	نتیجه‌گیری

### فصل چهارم رگرسیون و پیش‌بینی

۱۴۹	رگرسیون خطی ساده (Simple Linear Regression)
۱۵۲	رگرسیون خطی چندگانه (Multiple Linear Regression)
۱۵۴	رگرسیون ریبج (Ridge Regression)
۱۵۷	رگرسیون الاستیک نت (Elastic Net Regression)

۱۵۹.....	رگرسیون پولینومیال (Polynomial Regression)
۱۶۱.....	مدل‌های رگرسیون براساس فاکتور محبوبیت
۱۶۳.....	معیارهای بررسی عملکرد در مدل‌های رگرسیون
۱۶۳.....	کاربردهای متریک‌های مدل رگرسیون و انواع آن
۱۶۷.....	محبوبیت و کاربرد هر یک از متریک‌ها

## فصل پنجم طبقه‌بندی و پیش‌بینی ..... ۱۷۰

۱۷۰.....	رگرسیون لجستیک (Logistic Regression)
۱۷۴.....	الگوریتم K نزدیک‌ترین همسایه (K-Nearest Neighbors)
۱۸۱.....	درخت تصمیم (DTR (Decision Tree Model)
۱۸۴.....	مدل جنگل تصمیم (Random Forest)
۱۸۸.....	ماشین بردار پشتیبان (SVM)
۱۹۲.....	مدل SVR برای مسائل رگرسیون
۱۹۳.....	معیارهای ارزیابی مدل‌های طبقه‌بندی در یادگیری ماشین
۱۹۸.....	محبوبیت و کاربرد هر یک از معیارهای طبقه‌بندی

## فصل ششم الگوریتم‌های خوشه‌بندی ..... ۲۰۰

۲۰۲.....	الگوریتم K-Means
۲۰۶.....	خوشه‌بندی سلسله‌مراتبی (Hierarchical Clustering)
۲۱۰.....	الگوریتم DBSCAN (خوشه‌بندی فضایی مبتنی بر تراکم با نویز)
۲۱۵.....	الگوریتم OPTICS (شناسایی ساختار خوشه‌بندی با مرتب‌سازی نقاط)
۲۱۷.....	مدل پنهان مارکوف (HMM) چیست؟
۲۲۰.....	معیارهای ارزیابی دقت (Metrics) مدل‌های خوشه‌بندی

## فصل هفتم مدل‌های کاهش بعد (Dimensionality Reduction) ..... ۲۲۵

۲۲۸.....	مدل کاهش بعد PCA (تجزیه و تحلیل مؤلفه اصلی)
۲۳۱.....	مدل کاهش بعد ICA (تجزیه و تحلیل مؤلفه مستقل)
۲۳۸.....	مدل کاهش بعد RP (Random Projection)

## فصل هشتم مدل‌های دسته‌بندی (Ensemble Methods) ..... ۲۴۰

۲۴۵.....	مدل جنگل تصادفی (Random Forest) این بار برای دسته‌بندی
۲۴۹.....	مدل بوستینگ (Boosting)
۲۵۱.....	الگوریتم پایلینگ (Stacking)
۲۵۳.....	معیارهای متریک برای مدل‌های دسته‌بندی
۲۵۵.....	معیارهای متداول برای مدل‌های دسته‌بندی به ترتیب محبوبیت