

# متن کاوی نگرش یادگیری ماشین

مؤلف

چارو سی. آگاروال

مترجم

ایوب ترکیان

نیاز دانش

سرشناسه	: آگاروال، چارو سی، ۱۹۷۰ - م Aggarwal, Charu C
عنوان و نام پدیدآور	: متن کاوی نگرش یادگیری ماشین / مولف چارو سی آگاروال؛ مترجم ایوب ترکیان
مشخصات نشر	: تهران: نیاز دانش، ۱۳۹۷
مشخصات ظاهری	: ۶۷۶ ص: مصور، جدول، نمودار
شابک	: 978-600-8906-33-9
وضعیت فهرست‌نویسی	: فیپا
یادداشت	: عنوان اصلی: Machine learning for text, 2018.
موضوع	: فراگیری ماشینی Machine learning
موضوع	: متن پردازش (Computer science) Text processing
شناسه افزوده	: ترکیان، ایوب، ۱۳۳۷ - مترجم
رده‌بندی کنگره	: QC۳۲۵/۵/آ۷م۲ ۱۳۹۷
رده‌بندی دیویی	: ۰۰۶/۳۱
شماره کتابشناسی ملی	: ۵۴۲۹۶۴۳



نام کتاب	: متن کاوی نگرش یادگیری ماشین
مؤلف	: چارو سی آگاروال
مترجم	: ایوب ترکیان
مدیر اجرایی - ناظر بر چاپ	: حمیدرضا محمد شیرازی - محمد شمس
ناشر	: نیاز دانش
صفحه آرا	: واحد تولید انتشارات نیاز دانش
نوبت چاپ	: اول - ۱۳۹۷
شمارگان	: ۱۰۰ نسخه
قیمت	: ۶۵۰۰۰۰ ریال

ISBN:978-600-8906-33-9

شابک: ۹۷۸-۶۰۰-۸۹۰۶-۳۳-۹

هرگونه چاپ و تکثیر (اعم از زیراکس، بازنویسی، ضبط کامپیوتری و تهیه‌ی CD) از محتویات این اثر بدون اجازه کتبی ناشر ممنوع است، متخلفان به موجب بند ۵ از ماده ۲ قانون حمایت از مؤلفان، مصنفان و هنرمندان تحت پیگرد قانونی قرار می‌گیرند.

کلیه حقوق این اثر برای ناشر محفوظ است.

آدرس انتشارات: تهران، میدان انقلاب، خیابان ۱۲ فروردین، تقاطع وحید نظری، پلاک ۲۵۵، طبقه ۱، واحد ۲

۰۲۱-۶۶۴۷۸۱۰۶-۶۶۴۷۸۱۰۸-۰۹۱۲۷۰۷۳۹۳۵

[www.Niyez-Danesh.com](http://www.Niyez-Danesh.com)

مشاوره جهت نشر: ۲۱۰۶۷۰۹ - ۰۹۱۲

---

---

## فهرست مطالب

---

---

عنوان	شماره صفحه
فصل ۱ / مقدمه	۲۱
۱-۱ مقدمه	۲۱
۱-۱-۱ سازمان فصل	۲۴
۲-۱ اهمیت یادگیری از متن	۲۴
۳-۱ مدل‌های تحلیلی برای متن	۲۶
۱-۳-۱ پیش‌پردازش متن و محاسبه مشابهت	۲۷
۲-۳-۱ کاهش بُعد و فاکتورگیری ماتریس	۲۹
۳-۳-۱ خوشه‌بندی متن	۳۰
۱-۳-۳-۱ روش‌های قطعی و احتمالی فاکتورگیری ماتریس	۳۱
۲-۳-۳-۱ مدل‌های مخلوط احتمالی اسناد	۳۱
۳-۳-۳-۱ الگوریتم‌های مشابهت پایه	۳۲
۴-۳-۳-۱ روش‌های پیشرفته	۳۳
۴-۳-۱ طبقه‌بندی متن و مدل‌سازی رگرسیونی	۳۳
۱-۴-۳-۱ درخت‌های انتخاب	۳۵
۲-۴-۳-۱ طبقه‌گرهای قاعده‌پایه	۳۵
۳-۴-۳-۱ طبقه‌گر بیز ساده‌نگر	۳۵
۴-۴-۳-۱ طبقه‌گرهای نزدیک‌ترین همسایه	۳۶

۳۶	..... طبقه‌گرهای خطی ۵-۴-۳-۱
۳۷	..... موضوعات وسیع‌تر در طبقه‌بندی ۶-۴-۳-۱
۳۷	..... تحلیل مشترک متن با داده‌های نامتجانس ۵-۳-۱
۳۸	..... بازیابی اطلاعات و جستجوی وب ۶-۳-۱
۳۹	..... مدل‌سازی و تعبیه‌گری زبان متوالی ۷-۳-۱
۳۹	..... خلاصه‌سازی متن ۸-۳-۱
۳۹	..... استخراج اطلاعات ۹-۳-۱
۴۰	..... نظر کاوی و تحلیل تمایل ۱۰-۳-۱
۴۰	..... بخش‌بندی متن و ردیابی رویداد ۱۱-۳-۱
۴۰	..... خلاصه ۴-۱
۴۱	..... تمرین ۵-۱

## فصل ۲ / آماده‌سازی متن و محاسبه مشابهت ..... ۴۳

۴۳	..... مقدمه ۱-۲
۴۴	..... سازمان فصل ۱-۱-۲
۴۴	..... استخراج و توکن‌سازی متن خام ۲-۲
۴۸	..... مسایل خاص وب در استخراج متن ۱-۲-۲
۴۹	..... استخراج ترم از توکن ..... ۳-۲
۴۹	..... حذف واژه‌های توقف ۱-۳-۲
۵۰	..... خط تیره کوتاه ۲-۳-۲
۵۱	..... تغییر حالت حرف ۳-۳-۲
۵۱	..... هم‌سان‌سازی کاربرد پایه ۴-۳-۲
۵۲	..... ریشه‌یابی ۵-۳-۲
۵۳	..... نمایش فضای برداری و نرمال‌سازی ۴-۲
۵۵	..... محاسبه مشابهت در متن ۵-۲
۵۸	..... مفیدبودن نرمال‌سازی idf و ریشه‌یابی ۱-۵-۲
۵۹	..... خلاصه ۶-۲
۶۰	..... تمرین ۷-۲

۶۱	فصل ۳ / فاکتورگیری ماتریس و مدل سازی موضوعی
۶۱	۱-۳ مقدمه
۶۴	۱-۱-۳ سازمان فصل
۶۴	۲-۱-۳ نرمال سازی فاکتورگیری دوطرفه به سه طرفه
۶۶	۲-۳ تجزیه مقدار منفرد
۷۰	۱-۲-۳ مثال SVD
۷۲	۲-۲-۳ روش قدرت پیاده سازی SVD
۷۳	۳-۲-۳ کاربردهای SVD/LSA
۷۳	۴-۲-۳ محاسن و معایب SVD/LSA
۷۴	۳-۳ فاکتورگیری ماتریس غیرمنفی
۷۷	۱-۳-۳ تفسیرپذیری فاکتورگیری ماتریس غیرمنفی
۷۸	۲-۳-۳ مثال فاکتورگیری ماتریس غیرمنفی
۸۰	۳-۳-۳ کوتاه سازی در اسناد جدید
۸۱	۴-۳-۳ مزایا و معایب فاکتورگیری ماتریس غیرمنفی
۸۱	۴-۳ تحلیل سمانتیک ناپیدا احتمالی
۸۵	۱-۴-۳ ارتباط با فاکتورگیری ماتریس غیرمنفی
۸۶	۲-۴-۳ مقایسه با SVD
۸۶	۳-۴-۳ مثال PLSA
۸۷	۴-۴-۳ محاسن و معایب PLSA
۸۸	۵-۳ بررسی اجمالی تخصیص دیریکله نهفته
۸۸	۱-۵-۳ مدل LDA ساده شده
۹۲	۲-۵-۳ مدل LDA هموار شده
۹۳	۶-۳ تبدیل غیرخطی و مهندسی ویژگی
۹۷	۱-۶-۳ انتخاب تابع مشابهت
۹۷	۱-۱-۶-۳ توابع مشابهت کرنل سنتی
۱۰۰	۲-۱-۶-۳ تعمیم کیف واژه به N-گرام
۱۰۱	۳-۱-۶-۳ کرنل های زیرتوالی رشته
۱۰۵	۴-۱-۶-۳ تسریع زمان عملیات بازگشتی
۱۰۵	۵-۱-۶-۳ کرنل های تابع زبان

۱۰۶	تخمین نیستروم	۲-۶-۳
۱۰۸	موجودیت جزئی ماتریس مشابهت	۳-۶-۳
۱۱۱	خلاصه	۷-۳
۱۱۱	تمرین	۸-۳

## فصل ۴ / خوشه‌بندی متن

۱۱۵	مقدمه	۱-۴
۱۱۷	سازمان فصل	۱-۱-۴
۱۱۷	انتخاب و مهندسی ویژگی	۲-۴
۱۱۸	انتخاب ویژگی	۱-۲-۴
۱۱۸	قدرت ترم	۱-۱-۲-۴
۱۱۹	مدل‌سازی با نظارت برای انتخاب ویژگی بدون نظارت	۲-۱-۲-۴
۱۲۰	لفافی بدون نظارت با انتخاب ویژگی با نظارت	۳-۱-۲-۴
۱۲۰	مهندسی ویژگی	۲-۲-۴
۱۲۱	روش‌های فاکتورگیری ماتریس	۱-۲-۲-۴
۱۲۲	کاهش بعدیت غیرخطی	۲-۲-۲-۴
۱۲۳	تعبیه‌گری واژه	۳-۲-۲-۴
۱۲۳	مدل‌سازی موضوعی و فاکتورگیری ماتریس	۳-۴
۱۲۳	مدل‌های عضویت آمیخته و خوشه‌های هم‌پوشان	۱-۳-۴
۱۲۴	خوشه‌های بدون هم‌پوشانی و هم‌خوشه‌گی	۲-۳-۴
۱۲۸	هم‌خوشه‌گی با بخش‌بندی گراف دوجهتی	۱-۲-۳-۴
۱۲۹	مدل‌های آمیخته زایا برای خوشه‌بندی	۴-۴
۱۳۰	مدل برنولی	۱-۴-۴
۱۳۲	مدل مولتی‌نومی	۲-۴-۴
۱۳۴	مقایسه با مدل‌های موضوع با عضویت آمیخته	۳-۴-۴
۱۳۵	ارتباط با مدل بیز ساده‌شده برای طبقه‌بندی	۴-۴-۴
۱۳۶	الگوریتم k-میانگین	۵-۴
۱۳۹	هم‌گرایی و آغازگری	۱-۵-۴
۱۳۹	پیچیدگی محاسباتی	۲-۵-۴

۱۴۰	ارتباط با مدل‌های احتمالی	۳-۵-۴
۱۴۰	الگوریتم‌های خوشه‌بندی سلسله‌مراتبی	۶-۴
۱۴۴	پایاده‌سازی کارآمد و پیچیدگی محاسباتی	۱-۶-۴
۱۴۶	ارتباط طبیعی با $k$ - میانگین	۲-۶-۴
۱۴۷	دسته‌جمعی‌های خوشه‌بندی	۷-۴
۱۴۸	انتخاب مؤلفه دسته‌جمعی	۱-۷-۴
۱۴۹	تلفیق نتایج از مؤلفه‌های مختلف	۲-۷-۴
۱۴۹	خوشه‌بندی متن به صورت توالی	۸-۴
۱۵۰	روش‌های کرنلی برای خوشه‌بندی	۱-۸-۴
۱۵۱	$k$ - میانگین کرنلی	۱-۱-۸-۴
۱۵۲	مهندسی ویژگی صریح	۲-۱-۸-۴
۱۵۴	شگرد کرنل یا مهندسی ویژگی صریح؟	۳-۱-۸-۴
۱۵۵	کرنل‌های تابع داده: خوشه‌بندی طیفی	۲-۸-۴
۱۵۷	تبدیل خوشه‌بندی به یادگیری با نظارت	۹-۴
۱۵۸	موضوعات عملی	۱-۹-۴
۱۵۹	ارزیابی خوشه‌بندی	۱۰-۴
۱۵۹	چاله‌های سنجه‌های صحنه‌گذاری داخلی	۱-۱۰-۴
۱۶۰	سنجه‌های صحنه‌گذاری خارجی	۲-۱۰-۴
۱۶۴	رابطه ارزیابی خوشه‌بندی با یادگیری با نظارت	۱-۲-۱۰-۴
۱۶۵	اشتباهات رایج در ارزیابی	۲-۲-۱۰-۴
۱۶۶	خلاصه	۱۱-۴
۱۶۷	تمرین	۱۲-۴
۱۶۹	فصل ۵ / طبقه‌بندی متن: مدل‌های پایه	
۱۶۹	مقدمه	۱-۵
۱۷۱	انواع نشانه‌ها و مدل‌سازی رگرسیون	۱-۱-۵
۱۷۲	آموزش و تست	۲-۱-۵
۱۷۳	یادگیرنده‌های استقرایی، شرطی، و استنتاجی	۳-۱-۵
۱۷۴	مدل‌های بنیادی	۴-۱-۵

۱۷۵	۵-۱-۵	چالش‌های خاص متن در طبقه‌گرها
۱۷۵	۱-۵-۱-۵	سازمان فصل
۱۷۶	۲-۵	انتخاب و مهندسی ویژگی
۱۷۶	۱-۲-۵	نمایه Gini
۱۷۷	۲-۲-۵	انترپوی شرطی
۱۷۸	۳-۲-۵	اطلاعات جمعی جفت‌وار
۱۷۸	۴-۲-۵	سنج‌های با ارتباط نزدیک
۱۸۰	۵-۲-۵	آماره $\chi^2$
۱۸۱	۶-۲-۵	مدل‌های انتخاب ویژگی تعبیه‌گری
۱۸۲	۷-۲-۵	شگردهای مهندسی ویژگی
۱۸۳	۳-۵	مدل ساده‌شده بی‌ز
۱۸۴	۱-۳-۵	مدل برنولی
۱۸۵	۱-۱-۳-۵	فاز پیش‌بینی
۱۸۶	۲-۱-۳-۵	فاز آموزش
۱۸۷	۲-۳-۵	مدل مولتی‌نومی
۱۸۹	۳-۳-۵	مشاهدات عملی
۱۸۹	۴-۳-۵	درجه‌بندی خروجی‌ها با بی‌ز ساده‌شده
۱۹۰	۵-۳-۵	مثال بی‌ز ساده‌شده
۱۹۰	۱-۵-۳-۵	مدل برنولی
۱۹۲	۲-۵-۳-۵	مدل مولتی‌نومی
۱۹۴	۶-۳-۵	بی‌ز ساده‌شده نیمه‌نظارتی
۱۹۷	۴-۵	طبقه‌گر نزدیکترین همسایه
۱۹۸	۱-۴-۵	خصوصیات طبقه‌گر ۱-نزدیک‌ترین همسایه
۲۰۱	۲-۴-۵	طبقه‌بندی Ricchio و نزدیک‌ترین ثقل مرکزی
۲۰۳	۳-۴-۵	نزدیک‌ترین همسایه وزن‌دهی شده
۲۰۴	۱-۳-۴-۵	۱-نزدیک‌ترین همسایه کیسه‌ای و زیرنمونه
۲۰۶	۴-۴-۵	نزدیک‌ترین همسایه انطباقی: خانواده‌های قدرت‌مند
۲۰۹	۵-۵	درختان تصمیم و جنگل‌های راندومی
۲۱۰	۱-۵-۵	روش پایه برای ساخت درخت تصمیم



۲۱۲	..... تقسیم گره	۲-۵-۵
۲۱۲	..... پیش‌بینی	۱-۲-۵-۵
۲۱۳	..... تقسیم‌های چندمتغیره	۳-۵-۵
۲۱۳	..... موضوعات مشکل‌آفرین درخت تصمیم	۴-۵-۵
۲۱۵	..... جنگل‌های تصادفی	۵-۵-۵
۲۱۶	..... جنگل‌های تصادفی به‌عنوان نزدیک‌ترین همسایه انطباقی	۶-۵-۵
۲۱۷	..... طبقه‌گرهای قاعده‌پایه	۶-۵
۲۱۸	..... الگوریتم‌های با پوشش متوالی	۱-۶-۵
۲۱۹	..... قاعده یک-یادگیری	۱-۱-۶-۵
۲۲۱	..... هرس قاعده	۲-۱-۶-۵
۲۲۲	..... تولید قاعده از درخت تصمیم	۲-۶-۵
۲۲۳	..... طبقه‌گرهای انجمنی	۳-۶-۵
۲۲۴	..... پیش‌بینی	۴-۶-۵
۲۲۴	..... خلاصه	۷-۵
۲۲۵	..... تمرین	۸-۵

## فصل ۶ / طبقه‌بندی خطی و گرسیون برای متن ..... ۲۲۹

۲۲۹	..... مقدمه	۱-۶
۲۳۰	..... تفسیر هندسی مدل‌های خطی	۱-۱-۶
۲۳۲	..... متغیر سویه لازم است؟	۲-۱-۶
۲۳۳	..... تعریف کلی مدل‌های خطی با تنظیم	۳-۱-۶
۲۳۵	..... تعمیم پیش‌بینی‌های دوتایی به چندطبقه	۴-۱-۶
۲۳۶	..... خصوصیات مدل‌های خطی برای متن	۵-۱-۶
۲۳۷	..... نشانه‌گذاری فصل	۱-۵-۱-۶
۲۳۷	..... سازمان فصل	۲-۵-۱-۶
۲۳۷	..... رگرسیون کمترین مربعات و طبقه‌بندی	۲-۶
۲۳۸	..... رگرسیون کمترین مربعات با تنظیم $L_1$	۱-۲-۶
۲۳۹	..... پیاده‌سازی کارآمد	۱-۱-۲-۶
۲۳۹	..... تخمین تقریبی با تجزیه مقدار منفرد	۲-۱-۲-۶

- ۲۴۱-۳-۱-۲-۶ رابطه با رگرسیون مؤلفه‌های اصلی
- ۲۴۱-۴-۱-۲-۶ مسیر به رگرسیون کرنلی
- ۲۴۲-۲-۲-۶ LASSO: رگرسیون کمترین مربعات با تنظیم  $L_1$
- ۲۴۴-۱-۲-۲-۶ تفسیر LASSO به‌عنوان انتخاب‌گر ویژگی
- ۲۴۵-۳-۲-۶ تمایزگر خطی فیشر و طبقه‌بندی کمترین مربعات
- ۲۴۸-۱-۳-۲-۶ تمایزگر خطی با چند طبقه
- ۲۴۹-۲-۳-۲-۶ برابری تمایزگر فیشر و رگرسیون کمترین مربعات
- ۲۵۱-۳-۳-۲-۶ طبقه‌بندی کمترین مربعات تنظیم‌شده و LLSF
- ۲۵۳-۴-۳-۲-۶ پاشنه آشیل طبقه‌بندی کمترین مربعات
- ۲۵۴-۳-۶ ماشین‌های بردار پشتیبان
- ۲۵۵-۱-۳-۶ تفسیر بهینه‌سازی تنظیم‌شده
- ۲۵۶-۲-۳-۶ تفسیر حداکثر حاشیه
- ۲۵۸-۳-۳-۶ Pegasos: حل SVMها در اولی
- ۲۵۹-۱-۳-۳-۶ به‌روزرسانی‌های سازگار با تنکی
- ۲۶۰-۴-۳-۶ فرمولاسیون SVM دوگانه
- ۲۶۳-۵-۳-۶ الگوریتم‌های یادگیری برای SVMهای دوگانه
- ۲۶۴-۶-۳-۶ تفسیر نزدیکترین همسایه انطباقی برای SVMهای دوگانه
- ۲۶۶-۴-۶ رگرسیون لجیستیک
- ۲۶۷-۱-۴-۶ تفسیر بهینه‌سازی تنظیم‌شده
- ۲۶۹-۲-۴-۶ الگوریتم‌های آموزشی برای رگرسیون لجیستیک
- ۲۷۰-۳-۴-۶ تفسیر احتمالی رگرسیون لجیستیک
- ۲۷۱-۱-۳-۴-۶ تفسیر احتمالی گام‌های فرود گرادینتی استوکاستیک
- ۲۷۲-۲-۳-۴-۶ رابطه بین به‌روزرسانی‌های اولی مدل‌های خطی
- ۲۷۲-۴-۴-۶ رگرسیون لجیستیک مولتی‌نومی و دیگر تعمیم‌ها
- ۲۷۴-۵-۴-۶ توضیحات عملکرد رگرسیون لجیستیک
- ۲۷۵-۵-۶ تعمیم غیرخطی مدل‌های خطی
- ۲۷۷-۱-۵-۶ SVMهای کرنلی با تبدیل صریح
- ۲۷۸-۲-۵-۶ ترویج تفکیک‌پذیری خطی توسط کرنل‌های متعارف
- ۲۸۱-۳-۵-۶ قوت و ضعف کرنل‌های مختلف

۲۸۲	۱-۳-۵-۶	صید دانش زبان‌شناسی با کرنل‌ها
۲۸۲	۴-۵-۶	شگرد کرنلی
۲۸۳	۵-۵-۶	کاربرد سیستماتیک شگرد کرنلی
۲۸۹	۶-۶	خلاصه
۲۸۹	۷-۶	تمرین

## فصل ۷ / عملکرد و ارزیابی طبقه‌گر

۲۹۳	۱-۷	مقدمه
۲۹۴	۱-۱-۷	سازمان فصل
۲۹۴	۲-۷	تهاتر سویه-واریانس
۲۹۶	۱-۲-۷	نگاه رسمی
۳۰۱	۲-۲-۷	نشانه‌های گویای سویه و واریانس
۳۰۱	۳-۷	تبعات تهاتر سویه-واریانس بر عملکرد
۳۰۲	۱-۳-۷	اثر اندازه داده‌های آموزش
۳۰۴	۲-۳-۷	اثر بُعدیت داده‌ها
۳۰۵	۳-۳-۷	تبعات انتخاب مدل در متن
۳۰۶	۴-۷	بهبود عملکرد سیستماتیک با هنگردها
۳۰۶	۱-۴-۷	سبب‌سازی و زیرنمونه‌برداری
۳۰۸	۲-۴-۷	تقویت کردن
۳۱۱	۵-۷	ارزیابی طبقه‌گر
۳۱۳	۱-۵-۷	تقسیم به بخش‌های آموزش و تست
۳۱۳	۱-۱-۵-۷	بیرون نگه‌داشتن
۳۱۴	۲-۱-۵-۷	صحه‌گذاری تقاطعی
۳۱۴	۲-۵-۷	سنجه صحت مطلق
۳۱۴	۱-۲-۵-۷	صحت طبقه‌بندی
۳۱۶	۲-۲-۵-۷	صحت رگرسیون
۳۱۸	۳-۵-۷	سنجه‌های رتبه‌بندی برای طبقه‌بندی و بازیابی اطلاعات
۳۱۹	۱-۳-۵-۷	خصوصیات عملیاتی دریافت‌کننده
۳۲۴	۲-۳-۵-۷	قوی‌ترین سنجه‌ها برای لیست‌های درجه‌بندی

۳۲۶	..... خلاصه	۶-۷
۳۲۶	..... تمرین	۷-۷

## فصل ۸ / متن کاوی مشترک با داده‌های نامتجانس..... ۳۲۷

۳۲۷	..... مقدمه	۱-۸
۳۳۰	..... سازمان فصل	۱-۱-۸
۳۳۰	..... شگرد فاکتورگیری ماتریس اشتراکی	۲-۸
۳۳۱	..... گراف فاکتورگیری	۱-۲-۸
۳۳۲	..... کاربرد: فاکتورگیری اشتراکی با متن و لینک وب	۲-۲-۸
۳۳۴	..... حل مسئله به‌سازی	۱-۲-۲-۸
۳۳۶	..... تعبیه‌گری با نظارت	۲-۲-۲-۸
۳۳۷	..... کاربرد: متن با شبکه‌های اجتماعی بدون جهت	۳-۲-۸
۳۳۸	..... کاربرد به پیش‌بینی لینک با محتوای متن	۱-۳-۲-۸
۳۳۸	..... یادگیری انتقالی در تصاویر با متن	۴-۲-۸
۳۳۹	..... یادگیری انتقالی با متن بدون نشانه	۱-۲-۴-۸
۳۴۱	..... یادگیری انتقالی با متن نشانه‌دار	۲-۴-۲-۸
۳۴۳	..... کاربرد: سیستم‌های توصیه‌گر با درجه‌بندی و متن	۵-۲-۸
۳۴۵	..... کاربرد: متن کاوی بین‌زبانی	۶-۲-۸
۳۴۶	..... ماشین‌های فاکتورگیری	۳-۸
۳۵۱	..... روش‌های مدل‌سازی احتمالی مشترک	۴-۸
۳۵۲	..... مدل‌های احتمالی مشترک برای خوشه‌بندی	۱-۴-۸
۳۵۳	..... طبقه‌گر بیز ساده‌شده	۲-۴-۸
۳۵۳	..... تبدیل به روش‌های گراف کاوی	۵-۸
۳۵۶	..... خلاصه	۶-۸
۳۵۷	..... تمرین	۷-۸

## فصل ۹ / بازیابی اطلاعات و موتورهای جستجو..... ۳۵۹

۳۵۹	..... مقدمه	۱-۹
-----	-------------	-----

۳۶۰	.....	۱-۱-۹	سازمان فصل
۳۶۱	.....	۲-۹	نمایه‌سازی و پردازش پرسمان
۳۶۲	.....	۱-۲-۹	ساختارهای داده فرهنگ‌نامه
۳۶۵	.....	۲-۲-۹	نمایه وارونه
۳۶۶	.....	۳-۲-۹	ساخت نمایه زمان خطی
۳۶۹	.....	۴-۲-۹	پردازش پرسمان
۳۶۹	.....	۱-۴-۲-۹	بازیابی بولی
۳۷۰	.....	۲-۴-۲-۹	بازیابی درجه‌بندی‌شده
۳۷۲	.....	۳-۴-۲-۹	پردازش پرسمان با جمع‌زن ترم-در-هربار
۳۷۴	.....	۴-۴-۲-۹	پردازش پرسمان با جمع‌زن سند-در-هربار
۳۷۵	.....	۵-۴-۲-۹	ترم-در-هربار یا سند-در-هربار؟
۳۷۵	.....	۶-۴-۲-۹	انواع نمره رایج
۳۷۵	.....	۷-۴-۲-۹	پرسمان موقعیتی
۳۷۷	.....	۸-۴-۲-۹	نمره‌گذاری ناحیه‌ای
۳۷۸	.....	۹-۴-۲-۹	یادگیری ماشین در بازیابی اطلاعات
۳۸۱	.....	۱۰-۴-۲-۹	ماشین‌های بردار پشتیبان درجه‌بند
۳۸۳	.....	۵-۲-۹	به‌سازی‌های کارایی
۳۸۳	.....	۱-۵-۲-۹	نشانه‌گرهای گذری
۳۸۴	.....	۲-۵-۲-۹	لیست‌های قهرمان و نمایه‌های لایه‌ای
۳۸۵	.....	۳-۵-۲-۹	شگردهای ذخیره موقت
۳۸۶	.....	۴-۵-۲-۹	شگردهای فشرده‌سازی
۳۸۸	.....	۳-۹	نمره‌بندی با مدل‌های بازیابی اطلاعات
۳۸۹	.....	۱-۳-۹	مدل‌های فضای برداری با tf-idf
۳۹۰	.....	۲-۳-۹	مدل استقلال دوتایی
۳۹۳	.....	۳-۳-۹	مدل BM25 با فرکانس‌های ترم
۳۹۵	.....	۴-۳-۹	مدل‌های زبان آماری در بازیابی اطلاعات
۳۹۶	.....	۱-۴-۳-۹	مدل‌های شاییدی پرسمان
۳۹۸	.....	۴-۹	وب‌خیزی و کشف منبع
۳۹۹	.....	۱-۴-۹	یک الگوریتم خزنده پایه

۴۰۱	.....	۲-۴-۹	خزنده‌های ترجیحی
۴۰۳	.....	۳-۴-۹	چند رشته
۴۰۴	.....	۴-۴-۹	مقابله با تله‌های عنکبوت
۴۰۴	.....	۵-۴-۹	هم‌پوشی برای آشکارسازی تکرار نزدیک
۴۰۵	.....	۵-۹	پردازش پرسمان در موتورهای جستجو
۴۰۶	.....	۱-۵-۹	ساخت نمایه توزیعی
۴۰۷	.....	۲-۵-۹	به‌روزرسانی پویای نمایه
۴۰۷	.....	۳-۵-۹	پردازش پرسمان
۴۰۹	.....	۴-۵-۹	اهمیت اعتبار
۴۱۱	.....	۶-۹	الگوریتم‌های درجه‌بندی لینک‌پایه
۴۱۱	.....	۱-۶-۹	درجه‌صفحه
۴۱۵	.....	۱-۱-۶-۹	درجه‌صفحه حساس به موضوع
۴۱۶	.....	۲-۱-۶-۹	درجه‌بندی مشابهتی
۴۱۸	.....	۲-۶-۹	HITS
۴۲۰	.....	۷-۹	خلاصه
۴۲۰	.....	۸-۹	تمرین

## فصل ۱۰ / مدل‌سازی توالی متن و یادگیری عمیق ..... ۴۲۳

۴۲۳	.....	۱-۱۰	مقدمه
۴۲۶	.....	۱-۱-۱۰	سازمان فصل
۴۲۷	.....	۲-۱۰	مدل‌های زبان آماری
۴۳۱	.....	۱-۲-۱۰	مدل‌های گذرگرام
۴۳۲	.....	۲-۲-۱۰	رابطه با تعبیه‌گری‌ها
۴۳۴	.....	۳-۱۰	روش‌های کرنلی
۴۳۵	.....	۴-۱۰	مدل‌های فاکتورگیری ماتریس واژه-زمینه
۴۳۶	.....	۱-۴-۱۰	فاکتورگیری ماتریس با شمارش‌ها
۴۳۸	.....	۱-۱-۴-۱۰	سوژه‌های پس‌پردازش
۴۳۹	.....	۲-۴-۱۰	تعبیه‌گری GloVe
۴۴۰	.....	۳-۴-۱۰	فاکتورگیری ماتریس PPMI

- ۴۴۱-۴-۴-۱۰ فاکتورگیری ماتریس PPMI شیفت شده.....
- ۴۴۱-۵-۴-۱۰ گنجاندن نحو و دیگر ویژگی‌ها.....
- ۴۴۲-۵-۱۰ نمایش‌های گرافیکی فاصله واژه‌ها.....
- ۴۴۴-۶-۱۰ مدل‌های زبان عصبی.....
- ۴۴۴-۱-۶-۱۰ شبکه‌های عصبی: معرفی آرام.....
- ۴۴۵-۱-۱-۶-۱۰ لایه محاسباتی تکی: پرسپترون.....
- ۴۴۸-۲-۱-۶-۱۰ رابطه با ماشین‌های بردار پشتیبان.....
- ۴۵۰-۳-۱-۶-۱۰ انتخاب تابع فعال‌سازی.....
- ۴۵۱-۴-۱-۶-۱۰ انتخاب گره‌های خروجی.....
- ۴۵۱-۵-۱-۶-۱۰ انتخاب تابع اتلاف.....
- ۴۵۳-۶-۱-۶-۱۰ شبکه‌های عصبی چندلایه.....
- ۴۵۹-۲-۶-۱۰ تعبیه‌گری عصبی با Word2vec.....
- ۴۵۹-۱-۲-۶-۱۰ تعبیه‌گری عصبی با کیف‌واژه پیوسته.....
- ۴۶۳-۲-۲-۶-۱۰ تعبیه‌گری عصبی با مدل گذرگرام.....
- ۴۶۶-۳-۲-۶-۱۰ سوژه‌های عملی.....
- ۴۶۷-۴-۲-۶-۱۰ گذرگرام با نمونه‌برداری منفی.....
- ۴۶۹-۵-۲-۶-۱۰ معماری عصبی SGNS.....
- ۴۶۹-۳-۶-۱۰ Word2vec (SGNS) و فاکتورگیری ماتریس لجیستیک.....
- ۴۷۲-۱-۳-۶-۱۰ فرود گرادینانی.....
- ۴۷۳-۴-۶-۱۰ تعبیه‌گری پاراگراف با Doc2vec.....
- ۴۷۵-۷-۱۰ شبکه‌های عصبی برگشتی.....
- ۴۷۸-۱-۷-۱۰ سوژه‌های عملی.....
- ۴۷۸-۲-۷-۱۰ مثال مدلسازی زبان RNN.....
- ۴۸۰-۱-۲-۷-۱۰ تولید نمونه زبان.....
- ۴۸۱-۳-۷-۱۰ کاربرد به توضیح اتوماتیک تصویر.....
- ۴۸۳-۴-۷-۱۰ یادگیری توالی-به-توالی و ترجمه ماشین.....
- ۴۸۶-۱-۴-۷-۱۰ سامانه‌های سوال-جواب.....
- ۴۸۸-۵-۷-۱۰ کاربرد به طبقه‌بندی سطح جمله.....
- ۴۸۹-۶-۷-۱۰ طبقه‌بندی سطح توکن با ویژگی‌های زبان‌شناختی.....

۴۹۰	شبکه‌های برگشتی چندلایه	۷-۷-۱۰
۴۹۲	حافظه کوتاه مدت طولانی (LSTM)	۱-۷-۷-۱۰
۴۹۵	خلاصه	۸-۱۰
۴۹۶	تمرین	۹-۱۰

## فصل ۱۱ / خلاصه‌سازی متن ۴۹۹

۴۹۹	مقدمه	۱-۱۱
۵۰۰	خلاصه‌سازی استخراجی و چکیده‌ای	۱-۱-۱۱
۵۰۱	گام‌های کلیدی در خلاصه‌سازی استخراجی	۲-۱-۱۱
۵۰۲	فاز بخش‌بندی در خلاصه‌سازی استخراجی	۳-۱-۱۱
۵۰۲	سازمان فصل	۴-۱-۱۱
۵۰۲	روش‌های موضوع واژه و خلاصه‌سازی استخراجی	۲-۱۱
۵۰۳	احتمالات واژه	۱-۲-۱۱
۵۰۴	اوزان فرکانس نرمال شده	۲-۲-۱۱
۵۰۶	امضاءهای موضوع	۳-۲-۱۱
۵۰۸	روش‌های انتخاب جمله	۴-۲-۱۱
۵۰۹	روش‌های ناپیدا برای خلاصه‌سازی استخراجی	۳-۱۱
۵۰۹	تحلیل سمانتیک ناپیدا	۱-۳-۱۱
۵۱۱	زنجیره‌های واژگانی	۲-۳-۱۱
۵۱۱	شرح کوتاه WordNet	۱-۲-۳-۱۱
۵۱۲	بهره‌برداری از WordNet برای زنجیره‌های واژگانی	۲-۲-۳-۱۱
۵۱۴	روش‌های گراف پایه	۳-۳-۱۱
۵۱۵	خلاصه‌سازی ثقلی	۴-۳-۱۱
۵۱۶	یادگیری ماشین برای خلاصه‌سازی استخراجی	۴-۱۱
۵۱۷	استخراج ویژگی	۱-۴-۱۱
۵۱۸	طبقه‌گر مورد استفاده	۲-۴-۱۱
۵۱۸	خلاصه‌سازی چندسندی	۵-۱۱
۵۱۸	خلاصه‌سازی ثقل پایه	۱-۵-۱۱
۵۲۰	روش‌های گراف پایه	۲-۵-۱۱



۵۲۱	۶-۱۱ خلاصه‌سازی چکیده‌ای
۵۲۲	۱-۶-۱۱ فشرده‌سازی جمله
۵۲۳	۲-۶-۱۱ فیوژن اطلاعات
۵۲۳	۳-۶-۱۱ مرتب‌سازی اطلاعات
۵۲۴	۷-۱۱ خلاصه
۵۲۴	۸-۱۱ تمرین

## فصل ۱۲ / استخراج اطلاعات ۵۲۵

۵۲۵	۱-۱۲ مقدمه
۵۲۸	۱-۱-۱۲ تکامل تاریخی
۵۲۹	۲-۱-۱۲ نقش پردازش زبان طبیعی
۵۳۱	۳-۱-۱۲ سازمان فصل
۵۳۱	۲-۱۲ شناسایی مقوله نام‌دار
۵۳۳	۱-۲-۱۲ روش‌های قاعده‌پایه
۵۳۵	۱-۱-۲-۱۲ الگوریتم‌های آموزش برای سامانه‌های قاعده‌پایه
۵۳۶	۲-۱-۲-۱۲ تولید قاعده بالا به پایین
۵۳۷	۳-۱-۲-۱۲ تولید قاعده پایین به بالا
۵۳۸	۲-۲-۱۲ تبدیل به طبقه‌بندی سطح توکن
۵۳۹	۳-۲-۱۲ مدل‌های مارکوف پنهان
۵۴۰	۱-۳-۲-۱۲ مدل‌های مارکوف پنهان و مرئی
۵۴۰	۲-۳-۲-۱۲ سامانه Nymble
۵۴۲	۳-۳-۲-۱۲ آموزش
۵۴۳	۴-۳-۲-۱۲ پیش‌بینی برای بخش تست
۵۴۴	۵-۳-۲-۱۲ گنجاندن ویژگی‌های استخراج‌شده
۵۴۴	۶-۳-۲-۱۲ روایت‌ها و بهبودها
۵۴۵	۴-۲-۱۲ مدل‌های مارکوف انتروپی حداکثر
۵۴۷	۵-۲-۱۲ فیلدهای راندومی شرطی
۵۴۹	۳-۱۲ استخراج رابطه
۵۵۰	۱-۳-۱۲ تبدیل به طبقه‌بندی

۵۵۲	پیش‌بینی رابطه با مهندسی ویژگی صریح	۲-۳-۱۲
۵۵۳	استخراج ویژگی از توالی‌های جمله	۱-۲-۳-۱۲
۵۵۵	ساده‌سازی درختان تجزیه با گراف‌های وابستگی	۲-۲-۳-۱۲
۵۵۶	پیش‌بینی رابطه با مهندسی ویژگی: روش‌های کرنلی	۳-۳-۱۲
۵۵۷	کرنل‌ها از گراف‌های وابستگی	۱-۳-۳-۱۲
۵۵۸	کرنل‌های زیرتوالی‌پایه	۲-۳-۳-۱۲
۵۵۹	کرنل‌های درخت‌پایه کانولوشنی	۳-۳-۳-۱۲
۵۶۲	خلاصه	۴-۱۲
۵۶۲	تمرین	۵-۱۲

### فصل ۱۳ / نظرکاوی و تحلیل گرایش

۵۶۵	مقدمه	۱-۱۳
۵۶۸	واژگان نظر	۱-۱-۱۳
۵۶۹	رویکردهای واژه‌نامه‌پایه	۱-۱-۱-۱۳
۵۶۹	رویکردهای پیکره‌پایه	۲-۱-۱-۱۳
۵۷۰	نظرکاوی به‌عنوان سلول‌پرکنی و استخراج اطلاعات	۲-۱-۱۳
۵۷۲	سازمان فصل	۳-۱-۱۳
۵۷۲	طبقه‌بندی گرایش سطح سند	۲-۱۳
۵۷۴	رویکردهای بدون نظارت به طبقه‌بندی	۱-۲-۱۳
۵۷۶	طبقه‌بندی گرایش سطح جمله و سطح عبارت	۳-۱۳
۵۷۷	کاربردهای تحلیل سطح جمله و سطح عبارت	۱-۳-۱۳
۵۷۸	تبدیل طبقه‌بندی ذهنیت به مسئله قطع کمینه	۲-۳-۱۳
۵۷۸	زمینه در تحلیل قطبیت سطح جمله و عبارت	۳-۳-۱۳
۵۷۹	نظرکاوی جنبه‌پایه به‌عنوان استخراج اطلاعات	۴-۱۳
۵۸۰	رویکرد بدون نظارت Hu & Liu	۱-۴-۱۳
۵۸۲	OPINE: رویکرد بدون نظارت	۲-۴-۱۳
۵۸۳	استخراج نظر با نظارت به‌عنوان طبقه‌بندی سطح توکن	۳-۴-۱۳
۵۸۵	هرز نظر	۵-۱۳
۵۸۶	روش‌های با نظارت برای آشکارسازی هرز نظر	۱-۵-۱۳

۵۸۷	نشانه‌گذاری هرزنظر فریب‌کار	۱-۱-۵-۱۳
۵۸۸	استخراج ویژگی	۲-۱-۵-۱۳
۵۸۹	روش‌های بدون نظارت برای آشکارسازی هرزه‌گر	۲-۵-۱۳
۵۹۰	خلاصه‌سازی نظر	۶-۱۳
۵۹۰	خلاصه درجه‌بندی	۱-۶-۱۳
۵۹۱	خلاصه گرایش	۲-۶-۱۳
۵۹۱	خلاصه گرایش با عبارات و جملات	۳-۶-۱۳
۵۹۱	خلاصه‌های استخراجی و چکیده‌ای	۴-۶-۱۳
۵۹۲	خلاصه	۷-۱۳
۵۹۲	تمرین	۸-۱۳

## فصل ۱۴ / بخش‌بندی متن و آشکارسازی رویداد

۵۹۳	مقدمه	۱-۱۴
۵۹۴	رابطه با آشکارسازی و ردیابی موضوع	۱-۱-۱۴
۵۹۴	سازمان فصل	۲-۱-۱۴
۵۹۵	بخش‌بندی متن	۲-۱۴
۵۹۶	قطعه‌بندی متن	۱-۲-۱۴
۵۹۷	رویکرد C99	۲-۲-۱۴
۵۹۹	بخش‌بندی با نظارت با طبقه‌گرهای آماده	۳-۲-۱۴
۶۰۱	بخش‌بندی با نظارت با مدل‌های مارکوفی	۴-۲-۱۴
۶۰۴	کاوش جریان‌های متن	۳-۱۴
۶۰۴	خوشه‌بندی متن جاری	۱-۳-۱۴
۶۰۵	کاربرد به آشکارسازی اولین داستان	۲-۳-۱۴
۶۰۷	آشکارسازی رویداد	۴-۱۴
۶۰۸	آشکارسازی رویداد بدون نظارت	۱-۴-۱۴
۶۰۸	روش نزدیک‌ترین همسایه پنجره‌پایه	۱-۱-۴-۱۴
۶۰۹	بهره‌برداری از مدل‌های زایشی	۲-۱-۴-۱۴
۶۱۰	آشکارسازی رویداد در جریان‌های اجتماعی	۳-۱-۴-۱۴
۶۱۰	آشکارسازی رویداد و بخش‌بندی با نظارت	۲-۴-۱۴

---

۶۱۱	.....	۳-۴-۱۴	اشکارسازی رویداد و استخراج اطلاعات
۶۱۲	.....	۱-۳-۴-۱۴	تبدیل به طبقه‌بندی سطح توکن
۶۱۳	.....	۲-۳-۴-۱۴	استخراج رویداد زمینه باز
۶۱۵	.....	۵-۱۴	خلاصه
۶۱۵	.....	۶-۱۴	تمرین
۶۱۷	.....		پیوست «الف» - آموزش عملی متن کاوی
۶۶۹	.....		پیوست «ب» - فرهنگ‌نامه اصطلاحات