یادگیری ماشین تخاصمی

نگـرش امنیت اطلاعـات

مؤلفین

آنتونی جوزف

بلین نلسون

بنیامین روبینستاین

ج.د. تایگار

مترجم

ایوب ترکیان

نیاز دانش

فهرست مطالب

عنوان **شماره صفحه**

فصل ۱ / مقدمه ۷

۱.۱ انگیزه ۹
۲.۱ رویکرد قاعده‌مند به یادگیری امن ۱۸
۳.۱ سیر زمانی یادگیری امن ۲۱
۴.۱ نمای کلان ۲۱

فصل ۲ / پس‌زمینه و نشانه‌گذاری ۷۳

۱.۲ نشانه‌گذاری پایه ۲۵
۲.۲ یادگیری ماشین آماری ۲۶
۱.۲.۲ داده‌ها ۲۸
۲.۲.۲ فضای فرضیه ۳۰
۳.۲.۲ مدل یادگیری ۳۱
۴.۲.۲ یادگیری با نظارت ۳۱
۵.۲.۲ دیگر طرح‌واره‌های یادگیری ۳۵

فصل ۳ / چارچوب یادگیری امن ۱۱۳

۱.۳ تحلیل فازهای یادگیری ۳۸
۲.۳ تحلیل امنیت ۴۰
۱.۲.۳ اهداف امنیت ۴۰
۲.۲.۳ مدل تهدید ۴۱
۳.۲.۳ کاربردهای یادگیری ماشین در امنیت ۴۲
۳.۳ چارچوب ۴۳
۱.۳.۳ گروه‌بندی ۴۳
۲.۳.۳ بازی یادگیری تخاصمی ۴۵
۳.۳.۳ خصوصیات توانمندی‌های تخاصمی ۴۶
۴.۳.۳ حملات ۴۸
۵.۳.۳ دفاع‌ها ۴۹
۴.۳ حملات کاوشی ۴۹
۱.۴.۳ بازی کاوشی ۵۰
۲.۴.۳ حملات انسجام کاوشی ۵۱
۳.۴.۳ حملات قابلیت دسترسی کاوشی ۵۶
۴.۴.۳ دفاع مقابل حملات کاوشی ۵۷
۵.۳ حملات علّی ۶۳
۱.۵.۳ بازی علّی ۶۴
۲.۵.۳ حملات انسجام علّی ۶۵
۳.۵.۳ حملات در دسترس بودن علّی ۶۷
۴.۵.۳ دفاع مقابل حملات علّی ۶۹
۶.۳ بازی‌های یادگیری تکرارشده ۷۳

۱.۶.۳ بازی‌های یادگیری تکرارشده در امنیت ۷۶
۷.۳ یادگیری صیانت حریم خصوصی ۷۸
۱.۷.۳ حریم خصوصی دیفرانسیل ۷۸
۲.۷.۳ حملات حریم خصوصی کاوشی و علیّ ۸۰
۳.۷.۳ منفعت علیرغم راندوم بودن ۸۱

فصل ۴ / حمله به یادگیرنده فراکره‌ای ۱۶۵

۱.۴ حملات علیّ به آشکارسازهای فراکره‌ای ۸۵
۱.۱.۴ مفروضات یادگیری ۸۶
۲.۱.۴ مفروضات مهاجم ۸۷
۳.۱.۴ روش‌شناسی تحلیلی ۸۸
۲.۴ توصیف حمله فراکره‌ای ۸۹
۱.۲.۴ جابه‌جا کردن مرکز ثقل ۹۲
۲.۲.۴ توصیف ساختارمند حمله ۹۶
۳.۲.۴ ویژگی توالی‌های حمله ۹۸
۳.۴ حملات نامقید بهینه ۱۰۲
۱.۳.۴ پشته‌کردن بلوک‌ها ۱۰۳
۴.۴ تحمیل قیود زمانی روی حمله ۱۰۴
۱.۴.۴ پشته‌کردن بلوک‌های با جرم متغیر ۱۰۶
۲.۴.۴ فرمولاسیون آلترناتیو ۱۰۸
۳.۴.۴ راه‌حل رهاشده بهینه ۱۰۹
۵.۴ حمله علیه بازآموزی با جایگزینی داده‌ها ۱۱۳
۱.۵.۴ سیاست جایگزینی میانگین‌حذفی و راندوم‌حذفی ۱۱۵
۲.۵.۴ سیاست جایگزینی نزدیک‌حذفی ۱۱۷
۶.۴ مهاجمین مقید ۱۲۱
۱.۶.۴ حملات بهینه مقتصدانه ۱۲۳
۲.۶.۴ حملات با داده‌های مخلوط ۱۲۴
۳.۶.۴ بسط‌ها ۱۲۸
۷.۴ خلاصه ۱۲۹

فصل ۵ / مطالعه موردی حمله موجود بودن: SpamBayes ۲۱۷

۱.۵ فیلتر اسپم SpamBayes ۱۳۳
۱.۱.۵ الگوریتم آموزش ۱۳۳
۲.۱.۵ پیش‌بینی‌های SB ۱۳۵
۳.۱.۵ مدل SB ۱۳۶
۲.۵ مدل تهدید برای SpamBayes ۱۴۱
۱.۲.۵ اهداف مهاجم ۱۴۱
۲.۲.۵ دانش مهاجم ۱۴۲
۳.۲.۵ مدل آموزش دادن ۱۴۳
۴.۲.۵ فرض آلایش ۱۴۴
۳.۵ حملات علیّ علیه یادگیرنده SB ۱۴۵
۱.۳.۵ حملات موجود بودن علیّ ۱۴۵
۲.۳.۵ حملات انسجام علیّ– شبه‌اسپم ۱۵۰
۴.۵ دفاع ریجکت روی اثر منفی (RONI) ۱۵۰

۵.۵ آزمایشات با SpamBayes ۱۵۲
۱.۵.۵ روش تجربی ۱۵۲
۲.۵.۵ نتایج حمله فرهنگ‌نامه ۱۵۴
۳.۵.۵ نتایج حمله متمرکز ۱۵۶
۴.۵.۵ آزمایشات حمله شبه‌اسپم ۱۶۰
۵.۵.۵ نتایج RONI ۱۶۲
۶.۵ خلاصه ۱۶۴

فصل ۶ / مطالعه موردی حمله انسجام: آشکارساز PCA ۲۹۳

۱.۶ روش PCA آشکارسازی ناهنجاری‌های ترافیکی ۱۷۳
۱.۱.۶ ماتریس‌های ترافیک و ناهنجاری‌های حجم ۱۷۳
۲.۱.۶ روش زیرفضا برای آشکارسازی ناهنجاری ۱۷۵
۲.۶ ایجاد اختلال در زیرفضای PCA ۱۷۷
۱.۲.۶ مدل تهدید ۱۷۷
۲.۲.۶ انتخاب خاشاک ناآگاهانه ۱۷۹
۳.۲.۶ انتخاب خاشاک آگاهانه محلی ۱۷۹
۴.۲.۶ انتخاب خاشاک آگاهانه فراگیر ۱۷۹
۵.۲.۶ حملات قورباغه جوشان ۱۸۱
۳.۶ آشکارسازهای تاب‌آور در مقابل اختلال ۱۸۳
۱.۳.۶ حس درونی ۱۸۳
۲.۳.۶ PCA-GRID ۱۸۵
۳.۳.۶ حدآستانه لاپلاس ستبر ۱۸۷
۴.۶ ارزیابی تجربی ۱۹۰
۱.۴.۶ برپایش ۱۹۰
۲.۴.۶ شناسایی جریان‌های آسیب‌پذیر ۱۹۲
۳.۴.۶ ارزیابی حملات ۱۹۵
۴.۴.۶ ارزیابی ANTIDOTE ۱۹۷
۵.۴.۶ ارزیابی تجربی حمله مسموم‌سازی قورباغه جوشان ۱۹۹
۵.۶ خلاصه ۲۰۵

فصل ۷ / مکانیسم‌های حفظ حریم خصوصی یادگیری SVM ۳۴۳

۱.۷ مطالعات موردی رخنه به حریم ۲۰۷
۱.۱.۷ سوابق سلامت کارکنان دولتی ۲۰۸
۲.۱.۷ لاگ‌های پرسمان موتور جستجوی AOL ۲۰۸
۳.۱.۷ جایزه Netflix ۲۰۹
۴.۱.۷ کشف نام شبه‌نام‌های توییتر ۲۰۹
۵.۱.۷ مطالعات انجمنی سطح ژنوم ۲۱۰
۶.۱.۷ هدف‌گذاری تبلیغات میکرو ۲۱۰
۷.۱.۷ آموخته‌ها ۲۱۱
۲.۷ موقعیت مسئله: یادگیری حفظ حریم ۲۱۲
۱.۲.۷ حریم دیفرانسیل ۲۱۲
۲.۲.۷ منفعت ۲۱۵
۳.۲.۷ سوگیری تاریخی پژوهش در حریم دیفرانسیل ۲۱۶
۳.۷ ماشین‌های بردار پشتیبان: معرفی اجمالی ۲۱۹

۱.۳.۷ کرنل‌های تغییرناپذیر به جابه‌جایی ۲۲۱
۲.۳.۷ پایداری الگوریتمی ۲۲۱
۴.۷ حریم دیفرانسیل برحسب اختلال خروجی ۲۲۲
۵.۷ حریم دیفرانسیل برحسب اختلال هدف ۲۲۷
۶.۷ فضاهای ویژگی ابعاد نامعین ۲۳۰
۷.۷ مرزهای حریم دیفرانسیل بهینه ۲۳۹
۱.۷.۷ مرزهای بالایی ۲۳۹
۲.۷.۷ مرزهای پایینی ۲۴۲
۸.۷ خلاصه ۲۴۵

فصل ۸ / گریز نزدیک بهینه طبقه‌گرها ۳۴۳

۱.۸ توصیف گریز نزدیک بهینه ۲۵۱
۱.۱.۸ هزینه تخاصمی ۲۵۲
۲.۱.۸ گریز نزدیک بهینه ۲۵۴
۳.۱.۸ اصطلاحات جستجو ۲۵۶
۴.۱.۸ بهینه‌گی ضربی و جمع‌زدنی ۲۵۹
۵.۱.۸ خانواده طبقه‌گرهای القاگر تحدب ۲۶۲
۲.۸ گریز طبقات محدب برای هزینه‌های ℓ_1 ۲۶۴
۱.۲.۸ جستجوی $\epsilon - IMAC$ برای $\mathcal{X}_f^+$ محدب ۲۶۵
۲.۲.۸ یادگیری $\epsilon - IMAC$ برای $\mathcal{X}_f^-$ محدب ۲۷۷
۳.۸ گریز برای هزینه‌های ℓ_p کلی ۲۸۵
۱.۳.۸ مجموعه مثبت محدب ۲۸۵
۲.۳.۸ مجموعه منفی محدب ۲۹۲
۴.۸ خلاصه ۲۹۳
۱.۴.۸ مسایل باز در گریز نزدیک بهینه ۲۹۳
۲.۴.۸ معیارهای گریز آلترناتیو ۲۹۶
۳.۴.۸ گریز دنیای واقعی ۲۹۹

فصل ۹ / چالش‌های یادگیری ماشین تخاصمی ۳۴۳

۱.۹ بحث و مسایل باز ۳۰۹
۱.۱.۹ مؤلفه‌های بررسی نشده بازی تخاصمی و مسایل باز ۳۰۹
۲.۱.۹ توسعه فناوری‌های دفاعی ۳۱۱
۲.۹ نکات پایانی ۳۱۴

پیوست الف/ پس‌زمینه یادگیری و فراهندسه ۳۱۷
پیوست ب/ اثبات کامل حملات فراکره ۳۲۹
پیوست ج/ اثبات SpamBayes فصل ۵ ۳۳۹
پیوست د/ اثبات کامل گریز نزدیک بهینه ۳۴۹
پیوست ه/ متخاصم مبین ۳۵۹