بسمه تعالی

امنیت و حریم‌خصوصی در یادگیری ماشین

بخش گریز از دسته‌بند

۱- لطفا اسامی اعضای گروه‌های دو نفره خود و نام مقاله انتخابی‌تان را تا قبل از جلسه روز سه‌شنبه ۲۸ فروردین ماه در این صفحه وارد نمایید.

۲- افرادی که نتوانند تا تاریخ ذکر شده همگروهی پیدا کنند، در انتها جلسه روز سه‌شنبه ۲۸ فروردین ماه برای آن‌ها همگروهی انتخاب می‌شود.

۳- ارائه‌ها جزئی از درس هستند و در امتحان پایانی از آن‌ها سوال طرح می‌شود.

۴- نمره‌دهی ارائه‌ها توسط مدرسین و دانشجویان درس انجام می‌شود.

۵- برای ارائه یک پژوهش به موارد زیر دقت فرمایید.

۱. توضیح هدف مقاله
۲. توضیح راه‌حل ارائه شده
۳. بیان واضح ارزیابی
۴. بیان ضعف‌های پژوهش
۵. رابطه این پژوهش و پژهش‌های مرتبط

۶- هر ارائه به مدت ۲۰ دقیقه انجام می‌شود و ۱۰ دقیقه هم برای پرسش و پاسخ در نظر گرفته شده است.

۷- زمان و مکان برگزاری ارائه‌ها به زودی اطلاع‌رسانی می‌شود.

۸- اولویت تخصیص یک مقاله با گروهی است که زودتر آن را در صفحه ارائه‌ها وارد نماید. لطفا سطر مربوط به گروه‌های دیگر را ویرایش ننمایید و قبل از انتخاب مقاله بررسی کنید که آن مقاله تسط گروه دیگری انتخاب نشده باشد.

۹- در صورت سوال یا ابهام می‌تواند آن را در کوئرا یا به طور مستقیم با آدرس amsadeghzadeh@gmail.com مطرح نمایید.

۱۰- لیست مقالات نامزد برای ارائه در زیر آمده است. هر گروه باید یکی از مقالات را انتخاب نماید.

۱۱- تنها در صورت هماهنگی با استاد درس، امکان انتخاب مقاله‌ای خارج از لیست مقالات وجود دارد.

لیست مقالات نامزد

1. Audio Adversarial Examples: Targeted Attacks on Speech-to-Text
2. Perceptual Adversarial Robustness: Defense Against Unseen Threat Models
3. Adversarial Examples Are Not Bugs, They Are Features
4. Feature Purification: How Adversarial Training Performs Robust Deep Learning
5. On Adaptive Attacks to Adversarial Example Defenses
6. Fundamental Tradeoffs between Invariance and Sensitivity to Adversarial Perturbations
7. Adversarial Training for Free!
8. Reliable evaluation of adversarial robustness with an ensemble of diverse parameter-free attacks
9. Data Augmentation Can Improve Robustness
10. Square Attack: a query-efficient black-box adversarial attack via random search
11. Detecting Adversarial Examples Is (Nearly) As Hard As Classifying Them
12. Adversarial Examples for Malware Detection
13. Increasing Confidence in Adversarial Robustness Evaluations
14. On Improving Adversarial Transferability of Vision Transformers
15. Query-Efficient Hard-label Black-box Attack: An Optimization-based Approach
16. BAE: BERT-based Adversarial Examples for Text Classification