به نام ایزد یکتا

تمرین سوم درس روش پژوهش و ارائه

­­

استاد: دکتر رضا صفابخش

تهیه کننده: بردیا اردکانیان

9831072

**برسی و مهار مثال‌های خصمانه در یادگیری ماشین**

**سوال اول)**

**موضوع پروژه:** برسی و مهار مثال‌های خصمانه در یادگیری ماشین

برای یادداشت برداری بهتر از یک ساختار اولیه برای گزارش خود در نظر بگیریم. با برسی چکیده‌ها و مقدمه‌های منابع خود ساختار زیر را برای گزارش در نظر می‌گیریم.

1. مقدمه
2. مثال‌های خصمانه در یادگیری ماشین
   1. توضیح و تعریف مثال‌های خصمانه
   2. اهمیت مثال‌های خصمانه
3. کار‌های مرتبط
   1. برسی روش‌های ممکن برای پیدا کردن نمونه‌های متخاصم
   2. برسی طبقه بندی نمونه‌های متخاصم توسط أنواع طبقه‌بندی کننده‌ها
   3. برسی میزان درک یادگیری ماشین از مسئله مطرح شده
   4. برسی عملکرد یادگیری ماشین بر داده‌های طبیعی و داده‌هایی با احتمال توزع پایین
4. توضیح خطی مثال‌های خصمانه
5. اختلال خطی در مدل‌های غیرخطی
   1. روش fast gradient sign
6. آموزش خصمانه شبکه‌های عمیق
   1. برسی عملکرد شبکه‌های عمیق در برابر اغتشاش خصمانه
   2. قضیه universal approximator theorem
   3. آموزش ترکیبی از نمونه های متخاصم و پاک
7. چرا مثال های متخاصم تعمیم می‌دهند؟
8. نتیجه گیری

مراجع

طبیعی است این ساختار با جلو رفتن در مطالعات مقالات دستخوش تغییر شود

آ) مقاله [EXPLAINING AND HARNESSING ADVERSARIAL EXAMPLES](https://arxiv.org/pdf/1412.6572.pdf) نشانه گذاری شده است که در پیوست ضمیمه شده و مطالب مهم که قرار است در گزارش نوشتاری مورد استفاده قرار بگیرند highlight شده‌اند.

ب) از مقاله [EXPLAINING AND HARNESSING ADVERSARIAL EXAMPLES](https://arxiv.org/pdf/1412.6572.pdf) 17 فیش تهیه شده است که در فایل پاور پوینت ضمیمه شده است.