دانشگاه تهران

پردیس دانشکده‌های فنی

دانشکده برق و کامپیوتر

Trustworthy AI

تمرین شماره ۱

طراحان: عباس نصرت، رومینا اوجی

زمان تحویل: **۱۴۰۲/۰۱/۱۷**

زمستان ۱۴۰۱

فهرست

پرسش ۱ – Generalization and Robustness

در این تمرین قرار است با داده کم مدلی آموزش دهید که علاوه بر قدرت تعمیم (Generalization)، از مقاومت (Robustness) خوبی نیز برخوردار باشد.

۱- داده آموزش CIFAR10 را در نظر بگیرید. در این تمرین تنها ۲۰ درصد این داده را برای آموزش خود جدا می کنید و بقیه برای برازش مدل استفاده می شود. توجه داشته باشید که تناسب بین کلاس ها برقرار باشد.

۲- یک مدل ResNet18 را با استفاده از تابع هزینه CrossEntropy آموزش دهید. دقت آن را در دادگان برازش و تست گزارش کنید. خروجی قسمت کانولوشنال (backbone) شبکه را کاهش ابعاد داده و آن را برای داده دیده نشده نمایش دهید. (راهنمایی: می توانید از Umap برای کاهش ابعاد استفاده کنید. توجه داشته باشید که موقع آموزش umap برچسب کلاس ها را به آن ندهید. به عبارتی دیگر، umap را بصورت unsupervised اعمال کنید.)

۳- اغتشاشی در داده دیده نشده توسط مدل ایجاد کنید و مجددا دقت و تصویر خروجی backbone را بدست آورید. نتایج را با قسمت قبل مقایسه کنید. اغتشاش شما باید شامل آگمنتیشن هایی مانند نویز، color jitter و ... باشد. همچنین باید یک حمله متخاصمانه مانند fast gradient method به شبکه بزنید. (راهنمایی: پیشنهاد می شود از کتابخانه cleverhans برای حمله استفاده کنید.)

۴- قسمت ۲ و ۳ را با adversarial example ها مجددا تکرار کنید و نتایج را با حالت بدون Adversarial example مقایسه کنید. Adversarial example های شما باید شامل اغتشاشاتی که در قسمت ۳ اعمال کردید باشد.

۵- در مورد تابع هزینه AngularLoss توضیح دهید.

۶- بار دیگر قسمت ۲ و ۳ را برای تابع هزینه AngularLossانجام دهید و نتایج را با دو حالت قبل مقایسه کنید. (راهنمایی: توابع هزینه pytorch metric learning عمل ساخت تریپلت ها را بصورت خودکار انجام می دهند. تنها نکته ای که باید به آن توجه کنید این است که بچ شما باید شامل تعداد خوبی از کلاس ها باشد. برای این کار می توانید از BatchSampler استفاده کنید یا یک کلاس Dataset یا DataLoader کاستوم بنویسید.)

در صورتی که با کتابخانه TensorFlow راحت‌تر هستید، می توانید از کتابخانه TensorFlow Similarity استفاده کنید. توجه داشته باشید که در صورت نبودن هر کدام از توابع هزینه در این کتابخانه، باید آن را پیاده سازی کنید.

نکات پیاده سازی و تحویل

- مهلت ارسال این تمرین تا پایان روز "پنج‌شنبه ۱۷ فروردین ماه" خواهد بود.
- این زمان به هیچ وجه قابل تمدید نیست و درصورت نیاز میتوانید از grace time استفاده کنید.
- پیاده سازی با Tensorflow یا Pytorch باید انجام شود.
- انجام این تمرین به صورت یک نفره می‌باشد.
- در صورت مشاهده هر گونه تشابه در گزارش کار یا کدهای پیاده‌سازی، این امر به منزله تقلب برای طرفین در نظر گرفته خواهد شد.
- استفاده از کدهای آماده بدون ذکر منبع و بدون تغییر به منزله تقلب خواهد بود و نمره تمرین شما صفر در نظر گرفته می‌شود
- در صورت رعایت نکردن فرمت گزارش کار نمره گزارش به شما تعلق نخواهد گرفت.
- تمامی تصاویر و جداول مورد استفاده در گزارش کار باید دارای توضیح (caption) و شماره باشند.
- بخش زیادی از نمره شما مربوط به گزارش کار و روند حل مسئله است. خواهشا از هر گونه اطناب در گزارش کار پرهیز کرده و به موارد خواسته شده به صورت کامل پاسخ دهید.
- لطفا گزارش ، فایل کدها و سایر ضمائم مورد نیاز را با فرمت زیر در سامانه مدیریت دروس بارگذاری نمائید.

HW1_[Lastname]_[StudentNumber].zip

به طور مثال:

HW1_Nosrat_12345678.zip

- در صورت وجود سوال و یا ابهام میتوانید از طریق رایانامه زیر با موضوع HW1_TAI با دستیاران آموزشی در ارتباط باشید:

- پرسش اول

abbasnosrat@gmail.com یا تلگرام @a_b0ss

- تصحیح

romina.oji@ut.ac.ir

شاد و سلامت باشید.