به نام خدا

دانشکده مهندسی برق و کامپیوتر

درس یادگیری عمیق

تمرین سری اول

در این تمرین هدف پیاده‌سازی یک شبکه عصبی معمولی و یک Gaussian-RBF-Network و بررسی مقاومت آن‌ها بر روی حملات متخاصمانه است. مجموع داده‌های مورد استفاده در تمرین دیتاست fashion-MNIST می‌باشد. توجه داشته باشید که برای انجام این تمرین مجاز به استفاده از هیچ کد آماده و ابزارهای مرتبط با شبکه عصبی و یادگیری عمیق (نظیر pytorch و tensorflow) نمی‌باشید و باید تمامی مراحل بروزرسانی وزن‌ها و پارامترهای شبکه را خودتان پیاده‌سازی نمایید.

برای فهم بهتر و دقیق تر سوال پیشنهاد می شود مقاله زیر را مطالعه نمایید.

https://arxiv.org/pdf/1812.03190.pdf

تصاویر دیتاست fashion-MNIST تصاویر تک کانال و به سایز ۲۸*۲۸ هستند(معادل یک بردار ۷۸۴ بعدی) و از آنجا که به منظور کاهش زمان اجرا اندازه شبکه را کوچک تر در نظر می گیریم. در ابتدا به کمک PCA بعد داده ها را به ۱۲۸ تقلیل داده و سپس از آن ها به عنوان ورودی شبکه استفاده خواهیم نمود.

۱. ابتدا یک شبکه عصبی با یک لایه مخفی 150 نورونی و تابع فعال‌ساز relu و تابع خطا hinge_loss تعریف نمایید.در این مرحله باید پیاده‌سازی شما مستقل از تعداد لایه‌ها و تعداد نورون‌ها باشد به گونه ای که بتوانید به سادگی ابعاد شبکه را تغییر دهید.

۲. یک شبکه Gaussian-RBF-Network با لایه های مشابه ساختار سوال قبل و با تابع فعال ساز gelu و نرم L1 پیاده‌سازی نمایید که تابع Loss آن به صورت زیر است: ($\lambda = 550$)

$$J_{ML} = \sum_{i=1}^{N} \left(d_{y_i}(x^{(i)}) + \sum_{j \notin y_i} \max\left(0, \lambda - d_j(x^{(i)})\right) \right)$$

۳.شبکه‌های مطرح شده در قسمت‌های ۱و۲ را با استفاده از روش Stochastic Gradient Descent آموزش داده و نمودار تغییرات Loss و Accuracy را در طول یادگیری ترسیم نمایید. توجه کنید که مقدار دهی اولیه مناسب به وزن‌ها نقش مهمی در همگرایی شبکه ایفا می‌کند. ($batch\_size = 128$ و $Momentum = 0.9$)

۴. قسمت ۱ را برای حالتی که از PCA استفاده نمی کنیم و تمامی ۷۸۴ بعد را به عنوان ورودی در نظر می گیریم با استفاده از روش SGD آموزش دهید. در این حالت مقدار Loss و Accuracy و مدت زمان آموزش و تست را با قسمت ۱ مقایسه نمایید.

قسمت های امتیازی:

۵. اضافه کردن نویز به تصاویر ورودی شبکه در هنگام آموزش شبکه Gaussian-RBF-Network می تواند قابلیت اطمینان آن را افزایش دهد. با استفاده از additive isotropic Gaussian noise در ورودی، ساختار Gaussian-RBF-Network قبل را به یک شبکه de-noising تغییر دهید. ($\sigma = 0.2$). (توجه کنید که نویز به بردار ۱۲۸ بعدی پس از اعمال PCA اضافه شود).

بررسی مقاومت در برابر حملات متخاصنه.

ابتدا ده تصویر اول از داده های تست fashion-Mnist را جدا نموده و سایر مراحل را تنها برای این داده ها اجرا نمایید.

۶. برای هر داده ورودی و سپس برای هر ورودی تمامی نه کلاس اشتباه را به عنوان هدف حمله در نظر گرفته و با اعمال متد FGSM سعی کنید هر سه شبکه پیاده سازی شده در سوال های ۱ تا ۳ فریب دهید و درصد موفقیت برای هر شبکه را گزارش نمایید. موفقیت بدین معنا در نظر گرفته می شود که برای هر داده حداقل یکی از کلاس ها حمله موفقیت آمیز بوده باشد. در این قسمت برای پیاده سازی متد FGSM از توابع آماده استفاده نمایید.

۷. ۲۰۰ عدد داده نویز تصادفی(Isotropic Gaussian Noise) به مجموعه داده های آموزش شبکه اضافه نموده و برچسب آن‌ها را منفی یک در نظر بگیرد. ساختار مشابه سوال ۲ را با فرض ورود داده های rejection آموزش دهید. برای داده های با برچسب منفی یک تابع لاس به صورت زیر تعریف می شود.

$$\sum_{j} \max\left(0, \lambda - d_j(x^{(i)})\right)$$

بررسی threshold در مقاومت در برابر حملات متخاصمانه.

۸. شبکه سوال 7 را با سه مقدار متفاوت برای لاندا آموزش داده و مشابه سوال 6 موفقیت آمیز بودن حمله با تغییر threshold را بررسی کنید. همچنین برای هر کدام از مقادیر لاندا دقت شبکه بر روی داده های تست (accuracy) را گزارش نمایید.

نکات:

- در صورت مشاهده هر گونه مشابهت کد بین هر دو دانشجو، نمره تمرین هر دو دانشجو صفر لحاظ خواهد شد.
- در صورت مشاهده هر گونه مشابهت کد با کد های موجود در صفحات اینترنتی، نمره تمرین صفر لحاظ خواهد شد. اگر بخشی از کد را از کد آماده اینترنتی استفاده می کنید که جزو قسمت های اصلی تمرین نمی باشد، حتما باید لینک آن در گزارش و کد ارجاع داده شود.
- توجه نمایید که نیمی از نمره تمرین مربوط به گزارش می‌باشد. لازم به ذکر است رعایت اصول نگارشی حائز اهمیت است.
- در نوشتن گزارش، لحاظ جزییات نوشتن گزارش الزامی است. مانند موارد زیر:
  - ارجاع دادن به مطالب و اشکالی که از مقاله و وبسایت ها گرفته شده است.
  - توضیح اشکال و جداول در caption
  - نوشتن فرمول و قرار ندادن عکس مربوط به فرمول
  - ارجاع به شکل و جدول در متن گزارش
  - نوشتن نتایج شبیه سازی ها به صورت جدولی و شکل (از قرار دادن عکس نتیجه اجرای کد پرهیز شود)
  - درست بودن متن از نظر قواعد دستور زبانی و نگارشی
  - موارد تکمیلی در فایل template توضیح داده شده اند.
- گزارش تمرین را حتما به صورت PDF و در کنار کدهای تمرین در سایت درس آپلود نمائید.
- نحوه نامگذاری به صورت studentnumber_homeworknumber.pdf می باشد.
- زبان پیاده سازی python بوده و در این تمرین تنها مجاز به استفاده از numpy برای پیاده سازی شبکه هستید.
- برای پیاده سازی می توانید از محیط colab استفاده نمایید.
- هرگونه پرسش پیرامون تمرین را با ایمیل های aliparchekan@gmail.com و denial.ghiaseddin@ut.ac.ir مطرح نمایید.

موفق باشید.