



باسمه تعالی

دانشگاه صنعتی شریف

دانشکده مهندسی برق

۲۵۶۴۷-یادگیری عمیق-ترم پاییز

۱۴۰۰

تمرین سری چهارم

موعد تحویل:

نحوه تحویل:

- آپلود در CW در قالب یک فایل واحد با نام HW_04_stdnum.zip که stdnum شماره دانشجویی شما در دانشگاه صنعتی شریف می‌باشد.
- کدهای سوالات کامپیوتری را به صورت فایل‌های جداگانه و در فرمت ipynb تحویل دهید.
- دربرخی از قسمت‌ها نیاز به استفاده از تنسوربرد بوده، از این رو فایل Summary خود را حتماً تحویل دهید.
- از آنجایی که در طراحی‌های خود، مقید به محدودیتی نیستید، که چنین رویکردی منجر به خروجی‌های متفاوتی در همه ابعاد خواهد بود، از این رو مشاهده تقلب در تمارین باعث از دست‌دادن کل یا بخشی از نمره تمرین هر دو طرف می‌شود.
- در تکالیف شبیه‌سازی سهم عمده نمره تکلیف را تحلیل و دریافت شما از نتایج کدهای نوشته شده، دارد.
- از اجرای کدهای خود اطمینان حاصل فرمایید، ترجیحاً برای هماهنگی از version پکیج‌های مورد استفاده از فایل requirements.txt پکیج منیجر pip استفاده کنید. برای اطلاعات بیشتر به PEP ۴۹۶ یا PEP ۵۰۸ مراجعه کنید.
- برای تمارین کامپیوتری، استفاده از Git نمره‌ی امتیازی دارد، سعی کنید حتماً تعدادی Branch، Commit و Merge در Version Control خود داشته باشید. Branch‌های Merge شده را Remove نکنید. به نکاتی مانند قراردادن تصاویر و فایل‌هایی که به سورس کدتان مربوط نمی‌شود نیز دقت کنید.

۱. شبکه‌های تخصصی مولد

در طراحی تابع هدف GAN اگر G شبکه مولد و D شبکه تمیزدهنده باشد، در حالت پایه، تابع هدف به صورت زیر تعریف می‌گردد:

$$\mathcal{V}(G, D) = \mathbb{E}_{x \sim P_{Data}} [\log D(x)] + \mathbb{E}_{x \sim P_G} [\log(1 - D(x))]$$

که تابع مذکور به ازای G مینیمم و به ازای D به حد ماکسیمم خود میرسد.

الف) مقدار تمیزدهنده برحسب توزیع اصلی و مولد در حالتی که تمیزدهنده مستقر در شبکه با ظرفیت بالا، به حالت بهینه خود رسیده باشد، به چه صورت است؟

ب) چنانچه گام بهینه‌سازی نسبت به شبکه تمیزدهنده، به صورت بهینه صورت گیرد، نشان دهید کمینه کردن GAN با D بهینه، در گروهی کمینه کردن فاصله $JS(P_{Data} || P_G)$ است.

۲. مشکلات شبکه‌های تخصصی مولد

الف) Mode Collapse: این مشکل موجب می‌گردد تنوع نمونه‌های خروجی تولیدشده توسط شبکه کم شود. ابتدا نحوه استفاده از راهکار تولید مجموعه نمونه به جای تک‌نمونه جهت حل این مشکل، را شرح دهید؛ سپس اثر این راهکار بر فواصل KL JS و تغییرات تابع هزینه، بررسی نمایید.

ب) Non-convergence: یکی دیگر از مشکلات GAN امکان به وجود آمدن ناپایداری در بهینه‌سازی مسئله $\min_x \max_y f(x, y)$ است. برای مثال در تابع $f(x, y) = xy$ پایداری روش نزول گرادیان را برای رسیدن به نقطه بهینه به صورت تئوری و با حل معادلات دیفرانسیل آن بررسی کنید.

ج) Diminished gradient:

- چنانچه شبکه تمیزدهنده بسیار آموزش شود و نمونه‌های تولیدی مولد را با دقت بسیار بالا تشخیص دهد، گرادیانی که به مولد می‌رسد نزدیک به صفر است. تحقیق کنید چه تابع هزینه‌ای جهت حل این مشکل پیشنهاد شده است؟
- یکی دیگر از روش‌های حل این مشکل، اضافه نمودن نویز به ورودی تمیزدهنده است. این حالت مشابه آن است که هدف را کم کردن فاصله بین توزیع‌های نویزی شده مدل و داده اصلی مطرح نماییم. به صورت تئوری بررسی نمایید این روش به چه صورت گرادیان را بهبود می‌دهد؟

۳. امتیازی

در کنار دو فاصله‌ای که تاکنون آموخته‌اید، فاصله دیگری نیز به نام فاصله واسرشتاین وجود دارد که از محبوبیت زیادی در شبکه‌های تخصصی مولد برخورداراند.

فاصله واسرشتاین بین دو توزیع، کمترین هزینه جابه‌جایی یک توزیع به دیگری است. به صورت ریاضیاتی، فاصله واسرشتاین W_1 بین دو توزیع $q(x)$ و $p(x)$ به صورت زیر تعریف می‌گردد:

$$T(\gamma, c) = \int c(x, y) \gamma(x, y) dx dy$$

$$W_1(p, q) = \inf_{\gamma \in \Gamma(p, q)} T(\gamma, c)$$

که در آن، $c(x, y)$ هزینه جابه‌جایی و مقدار تابع $\gamma(x, y)$ نشان‌دهنده مقداری است که میان دو توزیع جابه‌جا می‌شود که در نتیجه $T(\gamma, c)$ هزینه جابه‌جایی بین توزیع‌های p و q تحت روش جابه‌جایی $\gamma(x, y)$ است.

الف) دو جفت توزیع مثال زده و سپس فاصله Wasserstein_1 و JS را میان هر جفت توزیع محاسبه نمایید.

ب) تحقیق نمایید چه فواصلی می‌توانند نشان‌دهنده یک روش جابه‌جایی باشند؟ سپس فرض کنید هزینه جابه‌جایی به صورت $c(x, y) = |x - y|$ باشد یک روش جابه‌جایی پیشنهاد دهید.

ج) با توجه به دو قسمت بالا، مزایای این فاصله نسبت به دو فاصله دیگر را بررسی نمایید.

DRAFT

هدف از این تمرین ابداع موسیقی با استفاده از شبکه ی GAN می باشد.

۱. از لینک دیتاست شامل ۸۰۰۰ قطعه موسیقی از ژانر های مختلف را دانلود کنید. در فایل Music.py با استفاده از تابع get_musics_coded ابتدا نويز pink موسیقی ها حذف می شوند و سپس تبدیل Mel Spectrogram محاسبه می شود. همانطور که می دانیم با استفاده از محاسبه ی FFT بر روی پنجره های همپوشاننده از سیگنال Spectrogram محاسبه می شود. اما انسان ها قادر به درک طیف محدودی از فرکانس و دامنه امواج صوتی هستند و همچنین درک تفاوت فرکانس های بالا و پایین برای انسان به یک اندازه نیست، برای مثال شاید تشخیص بین دو فرکانس ۵۰۰ و ۱۰۰۰ هرتز برای انسان راحت باشد ولی فرکانس های ۹۰۰۰ و ۹۵۰۰ هرتز به این شکل نباشد. برای حل این مشکل مقیاس مل (Mel Scale) معرفی شده است تا فرکانس ها از نظر شنونده فاصله های تقریباً یکسانی داشته باشند. فرمول تبدیل فرکانس به مقیاس مل به صورت زیر می باشد.

$$m = 2595 \log_{10} \left(1 + \frac{f}{700} \right)$$

۲. در نهایت خروجی تبدیل یک ماتریس دو بعدی خواهد بود. چند نمونه از این ماتریس ها را رسم کنید.
۳. برای ابداع موسیقی می توان از دو رویکرد کلی برای طراحی قسمت Generative شبکه ی GAN استفاده کرد. ماتریس خروجی دو بعدی را به صورت یک بردار درآورد از شبکه های بازگشتی (برای مثال LSTM و یا GRU) استفاده کرد و یا هر خروجی را به صورت دو بعدی نگه داشت و از شبکه های کانولوشنی استفاده کرد.
۴. اگر در هنگام آموزش شبکه ی GAN شما دچار یکی از مشکل های ذکر شده در سوال ۲ قسمت تئوری شد، آن را گزارش کنید و دلیل نتیجه گیری اینکه کدام یکی از مشکل Mode-collapse، Diminished gradient، Non-convergence را دارد بیان کنید. همچنین سعی کنید این مشکل را بهبود دهید.