

بسمه تعالی

درس بینایی ماشین

پروژه پایانی - بخش دو

هدف - می‌خواهیم به‌کارگیری به مرور و توسعه یادگیری مدل VAE پردازیم.
دادگان: مجموعه دادگان چهره با/بدون لب‌خند.

بخش ۱ - در مدل‌های VAE، هدف اصلی یادگیری توزیع پنهان (Latent Distribution) است که تا حد امکان به توزیع واقعی داده‌ها نزدیک باشد. از آنجا که محاسبه مستقیم درست‌نمایی لگاریتمی داده‌ها یعنی $\log(p(x))$ نیاز به محاسبات سنگینی دارد، از «حد پایین شواهد» طبق رابطه ELBO استفاده می‌شود:

$$ELBO = E_{q(z|x)} [\log p(x|z)] - D_{KL} (q(z|x) || p(z))$$

در اینجا $E_{q(z|x)} [\log p(x|z)]$ نشان‌دهنده دقت در بازسازی داده‌ها است. D_{KL} فاصله کولبک-لیبلر بین توزیع تقریبی $q(z|x)$ و توزیع پیش‌فرض پنهان $p(z)$ است.

با بیشینه‌سازی ELBO، بازسازی داده‌ها با افزایش جمله اول، بهبود یافته و نیز توزیع پنهان تقریبی به توزیع واقعی داده‌ها نزدیک‌تر می‌شود.

با تحلیل ریاضی نشان دهید که ELBO به صورت یک حد پایین از احتمال درست‌نمایی داده‌ها یعنی $\log(p(x))$ است.

بخش ۲ - همانطور که در سوال قبل و در کلاس گفته شد، آموزش مدل VAE بر بیشینه‌سازی ELBO استوار بوده که نیازمند نمونه‌برداری از توزیع $q(z|x)$ است. این نمونه‌برداری فرآیند آموزش VAE را به یک فرآیند تصادفی تبدیل می‌کند که محاسبه گرادیان‌ها و انجام پس‌انتشار آن‌ها را با مشکل روبه‌رو می‌کند زیرا نمونه‌برداری از یک توزیع، مشتق‌پذیر نیست.

برای برطرف کردن این مشکل از چه شیوه‌ای استفاده می‌شود؟ این شیوه را با نوشتن روابط مرتبط، به تفصیل توضیح دهید.

بخش ۳ - پیاده‌سازی VAE

می‌خواهیم گام به گام یک شبکه VAE را پیاده‌سازی کرده و بر دادگان تصویری چهره آموزش دهیم.

الف - ابتدا تصاویر را از پوشه به اشتراک گذاشته شده دریافت کرده و پس از پیاده‌سازی کلاس Data loader، چند تصویر تصادفی از مجموعه دادگان را نمایش دهید.

ب - یک کلاس برای پیاده‌سازی مدل VAE تعریف کنید و از ساختار ذکر شده در جدول ۱، برای تعریف کدگذار و کدگشا استفاده کنید. در ادامه، reparameterization trick پیاده‌سازی کرده و برای استفاده از آن، ساختار (لایه‌های) مورد نیاز را به کدگذار و کدگشا اضافه کنید.

در ایجاد ساختار شبکه، پس از هر لایه از تابع Relu استفاده کرده و برای لایه‌های کانولوشنی هم از Batch normalization و dropout با مقدار دلخواه استفاده کنید.

جدول ۱: تعریف لایه‌های مورد استفاده در شبکه β -VAE

Encoder Layers	Decoder Layers
Conv2d(3,32)	Linear(4096, h_dim)
Conv2d(32,64)	Reshape()
Conv2d(64,64)	Conv2d(64,64)
Conv2d(64,64)	Conv2d(64,64)
Flatten()	Conv2d(64,32)
Linear(h_dim, 4096)	Conv2d(32,3)

پ- تابع هزینه مورد استفاده در VAE را پیاده‌سازی کنید و شبکه را بر دادگان چهره آموزش دهید. برای آموزش شبکه از مقادیر جدول ۲ استفاده کنید. همچنین، در حین آموزش، هر یک از بخش‌های KL و Reconstruction در تابع هزینه را برای تصاویر آموزش و اعتبار سنجی در یک نمودار رسم کنید.

جدول ۲: پارامترهای مورد استفاده در آموزش مدل

Parameter	Value
Image dimensions	(128, 128)
Batch size	128
Transformation	Normalize, to tensor
Train/test ratio	0.8/0.2
Optimizer	Adam
Initial Learning Rate	0.0005
Epochs	1000

ت- پس از آموزش شبکه، ۸ تصویر را ابتدا با استفاده از تصاویر واقعی (اعمال کدگشا بر خروجی بردار نهان تصویر واقعی که از کدگذار عبور کرده) و سپس با استفاده از نویز تصادفی (اعمال کدگشا بر بردار نهانی که با استفاده از نویز تصادفی ساخته شده است) بازسازی کرده و در گزارش خود نمایش دهید.

ث- بعدها در فضای نهان معنادار بوده و هر یک بیانگر یکی از ویژگی‌های توزیع بدست آمده است. با توجه به اینکه دادگان مورد استفاده در آموزش مدل تنها تصاویر چهره انسان بوده و فقط شامل دو کلاس smile و non-smile بوده است، انتظار می‌رود که هر یک از بدهای بردار نهان مرتبط با یکی از اجزا چهره (لب، چشم‌ها و ...) باشد. برای آشکارسازی چنین معنایی، ابتدا بردار میانگین تمام تصاویر آموزشی کلاس smile و non-smile را به صورت جدا در فضای نهان بدست آورید. با کسر بردار نهان کلاس smile از بردار نهان کلاس non-smile انتظار داریم که بیشترین اندازه تفاضل در بُعد مرتبط با ویژگی‌های لب رخ دهد.

حال یک تصویر واقعی از دادگان را انتخاب کرده و با کسر/جمع ضریب‌های متفاوت از بردار تفاضلی بدست آمده از/با بردار تصویر انتخاب شده در فضای نهان، ۴ تصویر با درصد لبخندهای متفاوت تولید کرده و نمایش دهید.

م. ازوجی