

باسمه تعالی



نام استاد: دکتر فاطمی زاده

موضوع: پاسخ قسمت تئوری سوال چهارم از تمارین سری چهارم

تهیه کننده: نوید فرمehنی فراهانی

زمستان ۱۴۰۲

۱- برتری VQ-VAE به VAE معمولی:

- ۱- مدل VQ برای آموزش دادن ساده است و در صورتی که داده‌ها واریانس زیادی داشته باشند، مشکلی نخواهد بود؛ زیرا خروجی یک حالت گسسته دارد و نمیتواند یک عبارتی بین ۲ کلاس باشد (مانند بین ۲ کلاس سگ و گربه) [1].
- ۲- در VAE ها، مشکل Posterior Collapse وجود داشت و بدان معناست که از فضای نهفته به درستی استفاده نمی‌شود و احتمال پسین با پیشین برابر است و در کدگشا، خروجی مناسب تولید شود؛ اما هنگامی که از VQ-VAE ها استفاده می‌شود، این مشکل وجود نخواهد داشت [1]. ضمناً، این مشکل، مجدد وجود خواهد داشت اگر و وقت اگر در فضای نهفته مفاهیم به صورت غیر قابل تفهیم باشند [2]؛ به عبارت دیگر، طبق مباحث کلاس درس استاد، در صورتی که این مشکل وجود داشته باشد، اگر و فقط اگر فضای ما Entangled خواهد بود و هر یک از عناصر فضای نهفته، به درستی قسمت معینی از خروجی را بیان خواهند کرد.
- ۳- (که مربوط به برداشت بنده است) در صورتی که از VQ-VAE ها استفاده کنیم، میتوانیم خروجی‌های با معنا تولید کنیم.
- ۴- همانطور که در کلاس درس مطرح شد، در صورتی که از VQ-VAE ها استفاده شود، می‌توان به طور موثری فضای نهفته را Span کرد؛ به عنوان مثال object های در تصویر میتوانند پیکسلها را span کنند؛ لذا می‌توان خروجی‌های بسیار متنوعی را تولید نمود و تعداد این تنوع ها برابر با K^{h*w} خواهد بود که K تعداد بردارهای کتاب کد (به عنوان مثال، این عدد در تمرین، در ابتدا ۳ بود و در قسمت ط، به ۵۰ میرسید) و h و w، آخرین feature متصل به دیکدر است که vector ها از آن قسمت به decoder انتقال داده می‌شدند (که به عنوان مثال، هر دوی این اعداد میتواند ۷ باشد).
- ۵- به علت ویژگی‌های مطرح شده شامل Dis entangled بودن VQ-VAE ها، می‌توان از آن‌ها برای تبدیل صوت یک شخص به شکل دیگر، بدون عوض شدن کلمات استفاده کرد [1].

مراجع

- [1] A. Van Den Oord and O. Vinyals, "Neural discrete representation learning," Advances in neural information processing systems, vol. 30, 2017.
- [2] J. Rocca, "Understanding variational autoencoders (VAEs)," *Data Science*, 2019.