

بسم الله الرحمن الرحيم

دانشگاه علم و صنعت ایران

زمستان ۱۳۹۹

پاسخ تمرین سری چهاردهم

مبانی بینایی کامپیوتر

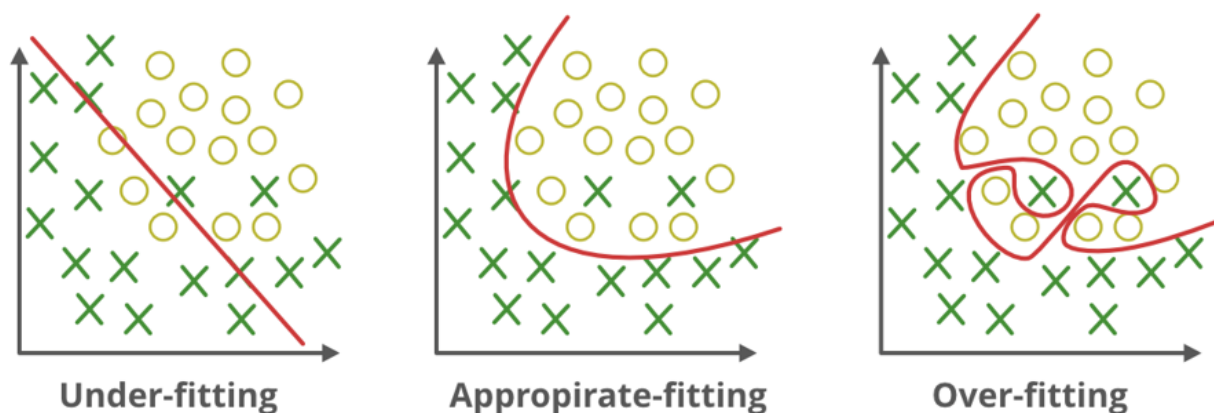
۱.

الف) علت به وجود آمدن مشکل `underfitting` در شبکه چیست و چه زمانی رخ می‌دهد؟

یک مدل یادگیری ماشین را زمانی می‌گوییم دچار `underfitting` شده است که نتواند رابطه میان داده‌های ورودی را پیدا کند. `underfitting` باعث می‌شود دقت، هم بر روی دادگان آموزشی و هم دادگان تست کم باشد. رخ دادن چنین مشکلی به زبان ساده به این معنی است که الگوریتمی که مدل به وسیله آن در حال یادگیری است، توانایی فیت شدن بر روی داده‌ها را ندارد. در این حالت قدرت تعمیم دهی مدل کم است و پیش‌بینی‌ها غیر قابل اعتماد است. مدل در این مواقع بسیار ساده است و یادگیری خاصی صورت نگرفته است. افزایش پیچیدگی مدل یا افزایش مدت زمان آموزش (افزایش تعداد Epochهای فرایند آموزش) از جمله راه‌های برطرف کردن مشکل `underfitting` است.

ب) مشکل بیش‌برازش (`overfitting`) چیست؟ چه راه‌هایی برای رفع آن پیشنهاد می‌کنید؟

یک مدل یادگیری ماشین زمانی می‌گوییم دچار `overfitting` شده است که به مقدار زیادی بر روی داده‌های آموزش، آموزش ببیند. این کار باعث می‌شود مدل جزئیات دارای نویز داده را یاد بگیرد، جزئیاتی که آن را در داده‌های آزمایشی دچار اشکال می‌کند. از دیگر علت‌های `overfitting` مدل، استفاده از مدل‌های پیچیده برای کارهای ساده است. این مدل‌ها به واسطه آزادی عملی که دارند، توانایی حفظ کردن پاسخ‌های داده‌های آموزشی را دارند. در این حالت مدل قابلیت تعمیم دهی ندارد. در این حالت دقت مدل روی دادگان آموزشی خوب است ولی بر روی دادگان تست دقت خوبی را به دست نمی‌آورد. استفاده از داده‌های آموزش بیشتر یا تکنیک‌هایی چون `data augmentation`، تکنیک‌های `regularization`، `early stopping`، `dropout` و ... از جمله راه‌های برطرف کردن `overfitting` است. همچنین ساده کردن شبکه می‌تواند مانع از بیش‌برازش شود.



۲. استفاده از شبکه‌های pre-train یکی از روش‌های موثر برای مقداردهی اولیه پارامترهای یک شبکه است. در مورد نحوه استفاده از این نوع یادگیری و مزایای آن، توضیح دهید.

یکی از بهترین روش‌ها برای مقداردهی اولیه پارامترهای یک شبکه استفاده از شبکه pre-train است. این روش که به آن اصطلاحاً انتقال یادگیری نیز می‌گویند در واقع به کارگیری دانش کسب شده در یک مسئله برای مسئله‌ای دیگر است. از این روش معمولاً هنگامی استفاده می‌شود که مسئله، داده‌های آموزشی کمی برای آموزش دیدن مدل در اختیار دارد. به عنوان مثال برای خیلی از کاربردهای پردازش تصویر از شبکه ResNet که توسط دیتاست ImageNet آموزش دیده است، استفاده می‌کنند.

برای آن که از روش انتقال دانش برای مقداردهی اولیه استفاده کنیم، بسته به سختی مسئله، مقدار داده‌ای که در اختیار داریم و میزان تفاوت مسئله با مسئله‌ای که مدل با آن آموزش دیده، یک یا چند لایه کاملاً متصل آخر مدل را حذف و لایه‌های جدید اضافه می‌کنیم. توجه کنید که بسته به تعداد کلاس‌های مسئله، ابعاد لایه خروجی تعیین می‌شود. سایر لایه‌ها در این روش بدون تغییر باید بمانند. به این کار اصطلاحاً freeze کردن لایه‌ها و به تغییر نامحسوس لایه‌های آخر مدل برای بهینه کردن آن برای مسئله دلخواه ما fine tune کردن می‌گویند. علت مناسب بودن این روش را می‌توان در یافتن نوع ویژگی‌های هر لایه جست و جو کرد. لایه‌های نزدیک به لایه ورودی، لایه‌هایی هستند که ویژگی‌های عمومی تصویر را مانند خطوط می‌یابند، در حالی که لایه‌های نزدیک‌تر به لایه‌های خروجی، ویژگی‌های پیچیده‌تر مانند یافتن اجزای صورت مانند چشم را می‌یابند. با نگهداشتن لایه‌های ابتدایی به نوعی ویژگی‌هایی که در همه مسائل نیاز هستند به صورت دقیق‌تری در مدل، مقداردهی اولیه می‌شوند.

از مزایای استفاده از شبکه‌های pre-train می‌توان به موارد زیر اشاره کرد:

- به راحتی قابلیت ترکیب شدن با یک مدل دیگر را دارند.
- در مدت زمان کمتری به دقت بالایی در مسئله می‌رسند.

- منابع بسیار کمتری (به عنوان مثال حافظه) در هنگام آموزش نیاز دارند.
- در زمان‌هایی که تعداد داده‌ها برای مسئله کم است، بسیار کارآمد هستند.

۳. اگر یک تصویر ۳ کاناله با ابعاد ۲۸ در ۲۸ داشته باشیم، تعداد پارامترها را برای یک لایه کانوولوشنی با تعداد ۱۲۸ فیلتر ۵×۵ را در حالتی که از کانوولوشن ساده و کانوولوشن با عمق جداپذیر (Depthwise Separable Convolution) استفاده می‌شود (به ازای depth_multiplier برابر با ۲)، به دست آورده و با هم مقایسه کنید.

استفاده از کانوولوشن عادی :

$$3 \times 5 \times 5 \times 128 + 128 = 9728$$

استفاده از کانوولوشن با عمق جداپذیر ($M = 2$) :

$$2 \times (5 \times 5 + 1) + 128 \times (3 \times 2 + 1) = 948$$

همانطور که مشاهده می‌شود به این خاطر که در کانوولوشن با عمق جداپذیر، کانوولوشن برای تک کانال‌ها انجام می‌شود و از فیلترهای ۱×۱ برای عمق استفاده می‌شود، تعداد پارامترها خیلی کمتر می‌شود (حدود ۰,۱ برابر).