

بسم الله الرحمن الرحيم

دانشگاه علم و صنعت ایران

زمستان ۱۳۹۹

پاسخ تمرین سری سیزدهم

مبانی بینایی کامپیوتر

۱.

a. فرض کنید یک حجم ۳۲ در ۳۲ در ۱۰ وارد یک لایه کانوولوشنی که ۱۶ فیلتر ۹ در ۹ دارد، می‌شود. مقدار گسترش مرزها چقدر باید باشد تا طول و عرض خروجی این لایه تفاوتی نکند. تعداد پارامترهای این لایه را نیز به دست آورید.

$$W_1 = 32, H_1 = 32, D_1 = 10$$

$$W_2 = 32, H_2 = 32, K = 16$$

$$W_2 = (W_1 - F + 2P)/S + 1$$

$$F = 9$$

اندازه گام را برابر یک در نظر می‌گیریم.

$$S = 1 \Rightarrow P = 4$$

پس مقدار گسترش مرزها از هر طرف ۴ خواهد بود.

تعداد پارامترها این لایه برابر است با:

$$F \cdot F \cdot D_1 \cdot K + K = 9 \times 9 \times 10 \times 16 + 16 = 12976$$

b. اگر یک تصویر ۳ کاناله با ابعاد ۳۲ در ۳۲ وارد یک لایه کانوولوشنی با ۳ فیلتر ۵ در ۵ بدون صفر افزونه و با اندازه گام ۱ شوند ابعاد خروجی چه خواهد شد؟ اگر همان تصویر را به دو لایه کانوولوشنی که هر دو ۳ فیلتر ۳ در ۳ بدون صفر افزونه و اندازه گام ۱ دارند بدهیم، ابعاد خروجی چه خواهد شد؟

اگر یک تصویر ۳ کاناله با ابعاد ۳۲ در ۳۲ وارد یک لایه کانوولوشنی با ۳ فیلتر ۵ در ۵ بدون صفر افزونه و اندازه گام ۱ شود، ابعاد خروجی ۲۸ در ۲۸ در ۳ خواهد شد.

حال اگر یک تصویر ۳ کاناله با ابعاد ۳۲ در ۳۲ وارد یک لایه کانولوشنی با ۳ فیلتر ۳ در ۳ بدون صفر افزونه و اندازه گام ۱ شود، ابعاد خروجی ۳۰ در ۳۰ خواهد شد. و این خروجی دوباره وارد لایه‌ای با همین ابعاد و تنظیمات شود، ابعاد خروجی ۲۸ در ۲۸ خواهد شد.

c. مطلع هستید که learning rate، batch size و epoch از سه پارامترهایی هستند که در شبکه‌های عمیق استفاده می‌شوند. تاثیر استفاده از هر کدام از این پارامترها را به صورت دقیق توضیح دهید.

تاثیر سه مورد را به ترتیب بررسی می‌کنیم:

Learning rate: در کل، سرعت آموزش شبکه را کنترل می‌کند که معمولاً بین ۰ تا ۱ قرار دارد. برای مثال اگر مقدارش برابر با ۰,۱ باشد به این معناست که هر سری وزن‌های شبکه به اندازه‌ی ۱۰ درصد خطا تخمین زده شده به‌روز می‌شوند. به طور کلی **learning rate** با میزان زیاد، سرعت یادگیری شبکه را افزایش می‌دهد ولی به دلیل این‌که با گام‌های بلندتری به‌روزرسانی می‌شود (درصد تغییر وزن بیشتر است) احتمال رسیدن به وزن‌های **sub optimal** وجود دارد. برعکس، با مقدار کم با اینکه سرعت یادگیری کاهش می‌یابد ولی وزن‌های یافت شده، وزن‌هایی نزدیک به وزن‌های **optimal** و یا حتی **global optimal** هستند. یک نقطه‌ی منفی برای آپدیت‌های بزرگ برای وزن‌ها می‌تواند **numerical overflow** باشد که با بزرگ شدن پی در پی وزن‌ها ممکن است صورت بگیرد.

Batch size: برابر با تعداد نمونه‌های استفاده شده از داده‌ی یادگیری در هنگام تخمین خطای گرادیان است و به طور کلی دقت خطای گرادیان تخمین زده شده با این هاپرپارامتر تعیین می‌شود زیرا هر چه تعداد نمونه‌ها بیشتر شود دقت بالاتر می‌رود. برای مثال اگر مقدارش برابر با ۳۲ باشد به این معناست که ۳۲ نمونه از داده‌ی آموزشی برای تخمین خطای گرادیان قبل از آپدیت کردن وزن‌ها استفاده می‌شود. اگر مقدار این هاپرپارامتر بیش از اندازه زیاد باشد میزان **generalization** و سرعت یادگیری را پایین می‌آورد.

Epoch: یک **epoch** یادگیری به این معناست که الگوریتم یادگیری یک بار روی داده‌ی آموزشی، که نمونه‌ها به طور رندوم در گروه‌هایی با اندازه‌ی **batch size** تقسیم شده‌اند، انجام می‌شود. میزان مناسب این هاپر پارامتر وابسته به دیتاست مورد استفاده است و به این صورت نیست که هر چه بیشتر باشد بهتر است. در یک نقطه‌ای افزایش بیشتر آن موجب **over fit** شدن می‌شود. یک مثال می‌تواند نمودار زیر باشد.

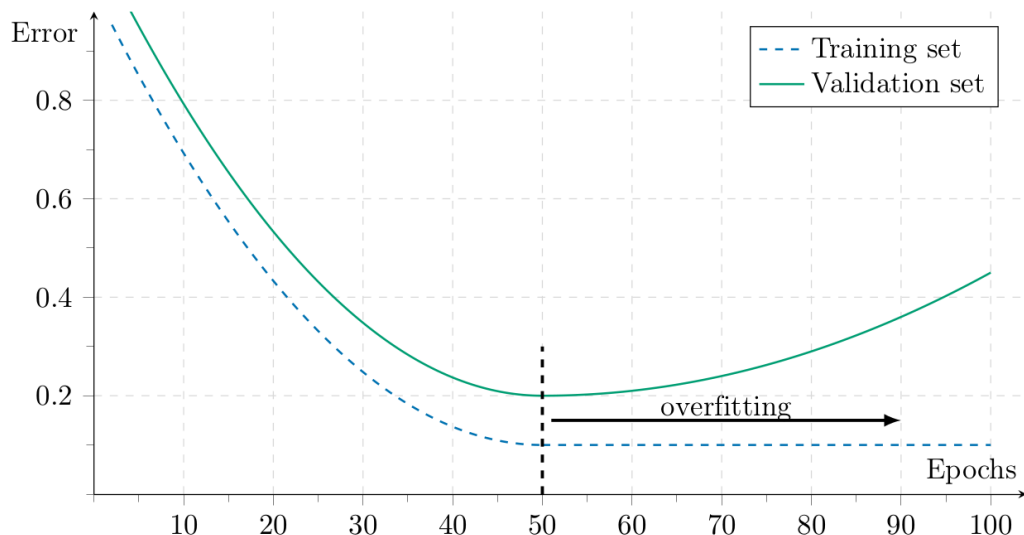


Figure 2.7.: A typical validation curve: In this case, the hyperparameter is the number of epochs and the quality metric is the error ($1 - \text{accuracy}$). The longer the network is trained, the better it gets on the training set. At some point the network is fit too well to the training data and loses its capability to generalize. At this point the quality curve of the training set and the validation set diverge. While the classifier is still improving on the training set, it gets worse on the validation and the test set.

d. مزایای استفاده از لایه‌های کانولوشنی نسبت به لایه‌های کاملاً متصل در پردازش تصویر چیست؟

شبکه عصبی کاملاً متصل خاصیت محلی بودن اطلاعات یک پیکسل و وجود و اهمیت همسایگی یک پیکسل را در نظر نمی‌گیرد ولی شبکه کانولوشنی به تصویر به عنوان یک تصویر با ابعاد و حجم نگاه می‌کند.

e. دلایل استفاده از لایه‌های ادغام (pooling) چیست؟

لایه ادغام باعث کم شدن حجم خروجی یک لایه کانولوشنی شده و پارامترهای مورد نیاز شبکه را کمتر می‌کند در حالی که اگر مناسب انتخاب شود اثر منفی بر روی فرآیند آموزش نمی‌گذارد و سرعت آن را نیز افزایش خواهد داد اما اگر به صورت مناسب انتخاب نشود ممکن است باعث از دست رفتن اطلاعات بشود.

f. نشان دهید که می‌توان به جای یک لایه کانولوشنی با فیلترهای $(7,7)$ از سه لایه متوالی با فیلترهای $(3,3)$ استفاده کرد و این کار محاسبات کمتری دارد.

در لایه کانولوشنی محاسباتی که داریم فقط ضرب و جمع هستند. به عنوان مثال فرض کنید یک تصویر 32 در 32 داریم. ابعاد خروجی بعد از اعمال یک لایه کانولوشنی با فیلتر 7 در 7 و

اندازه گام ۱ و بدون صفر افزونه ۲۶ در ۲۶ خواهد شد. تعداد ضرب های لازم برای محاسبه خروجی به این صورت است که برای محاسبه هر پیکسل از تصویر خروجی نیاز است که ۴۹ ضرب (چون تعداد اعداد فیلتر ۷ در ۷، ۴۹ عدد است) انجام شود. پس تعداد ضرب های لازم برابر $26 \times 26 \times 49 = 33124$ است.

حالا می خواهیم تعداد ضرب های لازم را در حالتی که از سه لایه کانوولوشنی با ابعاد فیلتر ۳ در ۳ استفاده می کنیم، محاسبه کنیم. فقط لازم است که در هر مرحله ابعاد خروجی را محاسبه کرده و این نکته را به خاطر داشته باشیم که برای محاسبه هر پیکسل از خروجی این بار به ۹ ضرب (چون تعداد اعداد فیلتر ۳ در ۳، ۹ عدد است) نیاز است. در نتیجه تعداد ضرب ها در این حالت به صورت زیر قابل محاسبه است:

$$(30 \times 30 \times 9) + (28 \times 28 \times 9) + (26 \times 26 \times 9) = 21240$$

قابل مشاهده است که ابعاد خروجی ثابت است ولی تعداد ضرب های لازم برای محاسبه خروجی کمتر شده است.

محاسبه برای تعداد جمع ها هم به همین صورت است با این تفاوت که برای هر پیکسل یک جمع بیشتر برای جمع با بایاس فیلتر نیز لازم است.

در نتیجه می توان به جای استفاده از یک فیلتر ۷ در ۷ کانوولوشنی از ۳ فیلتر ۳ در ۳ کانوولوشنی استفاده کرد چون ابعاد خروجی برابر است ولی تعداد محاسبات نیز کمتر است و حتی تعداد پارامترهایی که باید آموزش داده شود نیز کمتر خواهد شد. البته در مورد تعداد محاسبات این باید در نظر گرفته شود که ابعاد تصویر ورودی نیز باید نسبت به فیلترها بزرگ باشد تا در مورد حجم محاسبات چنین شرایطی که توضیح داده شد رخ دهد و ابعاد ورودی نیز قابل اهمیت است.