شبکه‌های عصبی مصنوعی
تکلیف سوم
روابط انتشار-به-عقب برای شبکه‌های کانولوشنی

داریوش حسن‌پور آده

۹۳۰۸۱۶۴

اگر نورون‌های ورودی به صورت یک مربع $N \times N$ در نظر بگیریم و همچنین هریک از نورون‌های کانولوشن به $m \times m$ عدد نورون‌های ورودی متصل باشند، اندازه‌ی لایه‌ی کانولشن برابر خواهد بود با مربعی با تعداد نورون $(N-m+1) \times (N-m+1)$ خروجی نورون‌های لایه‌ی کانولشن را به صورت ۲ می‌باشد.

$$x_{ji}^{l} = b + \sum_{l=0}^{m-1} \sum_{m=0}^{m-1} w_{l,m} \cdot a_{j+l,k+m} \tag{1}$$

$$y_{ji}^{l} = \sigma(x_{ji}^{l}) \tag{2}$$

در لایه‌ی Max-Pooling کار خاصی که نیاز به یادگیری داشته باشد انجام نمی‌دهیم(به این معنی که وزنی بین لایه‌ی کانولوشن و لایه‌ی Max-Pooling وجود ندارد) زیرا که فقط خروجی یک دسته‌ی $k \times k$ از لایه‌ی کانولشن را می‌گیرد و حداکثر مقدار آن‌ها را برمی‌گرداند. در نتیجه فقط نیاز داریم که برای وزن‌ها متصل به لایه‌ی خروجی از لایه‌ی Max-Pooling و وزن‌های متصل به لایه‌ی کانولوشن از لایه‌ی ورودی را بروز رسانی کنیم.

فرمول‌های انتشار-به-عقب و بروز رسانی وزن‌های متصل به لایه‌ی خروجی از لایه‌ی Max-Pooling همانند سابق می‌باشد و تغییری نکرده‌اند. بدین صورت که اگر تابع هزینه‌ی C را داشته باشیم روابط ۳ تا ۵ را خواهیم داشت. برای محاسبه‌ی انتشار-به-عقب برای وزن‌هایی که از لایه‌ی ورودی به لایه‌ی کانولوشن می‌آیند را با توجه به اینکه از تکنیک وزن‌های مشترک استفاده می‌کنیم باید به ازای هر وزن مقدار $\frac{\partial C}{\partial w_{l,m}}$ از آنجایی که در لایه‌ی کانولوشن $(N-m+1) \times (N-m+1)$ نورون دارد باید با استفاده از مشتق زنجیره‌ای میزان سهم گرادیانی هر وزن در خطای آن لایه را بدست بیاوریم، که در معادله‌ی ۷ آورده شده است.

$$\delta^l = ((w^{l+۱})^T \delta^{l+۱}) \odot \sigma'(z^l) \tag{۳}$$

$$\frac{\partial C}{\partial b_j^l} = \delta_j^l \tag{۴}$$

$$\frac{\partial C}{\partial w_{jk}^l} = a_k^{l-۱} \delta_j^l \tag{۵}$$

(۶)

$$\frac{\partial C}{\partial w_{l,m}} = \sum_{i=۰}^{N-m} \sum_{j=۰}^{N-m} \frac{\partial C}{\partial x_{ij}^l} \cdot \frac{\partial x_{ij}^l}{\partial w_{l,m}}$$

$$\overset{\text{eq.2}}{=} \sum_{i=۰}^{N-m} \sum_{j=۰}^{N-m} \frac{\partial C}{\partial x_{ij}^l} \cdot a_{i+l,j+m}^{l-۱} \tag{۷}$$

علت اینکه در ۷ از جمع استفاده شده است بخاطر این است که در لایه‌ی کانولوشن از تکنیک وزن‌های مشترک[۱] استفاده شده است. حال برای اینکه مقادیر گرادین را حساب کنیم باید مقادیر $\delta_i^l = \frac{\partial C}{\partial x_{ij}^l}$ را به بدست بیاوریم(رابطه‌ی ۸). همچنین برای بروز رسانی وزن‌ها در لایه‌ی کانولشن باید خطا را به لایه‌ی عقبی انتشار دهیم که رابطه‌ی ۹ خطای مرتبط به لایه‌ی پیشین را به ما می‌دهد را خواهیم داشت.

$$\frac{\partial C}{\partial x_{ij}^l} = \frac{\partial C}{\partial a_{ij}^l} \cdot \frac{\partial a_{ij}^l}{\partial x_{ij}^l} = \frac{\partial C}{\partial a_{ij}^l} \sigma'(x_{ij}^l) \tag{۸}$$

$$\frac{\partial C}{\partial a_{ij}^{l-۱}} = \sum_{i=۰}^{N-m} \sum_{j=۰}^{N-m} \frac{\partial C}{\partial x_{(i-l),(j-m)}^l} \cdot \frac{\partial x_{(i-l),(j-m)}^l}{\partial a_{ij}^{l-۱}}$$

$$\overset{\text{eq.2}}{=} \sum_{i=۰}^{N-m} \sum_{j=۰}^{N-m} \frac{\partial C}{\partial x_{(i-l),(j-m)}^l} \cdot w_{l,m} \tag{۹}$$

در نهایت وزن‌ها را مطابق معمول بروز رسانی خواهیم کرد.

$$w_{ij}^l = w_{ij}^l - \alpha \frac{\partial C}{\partial w_{ij}^l} \tag{۱۰}$$

[۱] Weight Sharing