

شبکه‌های عصبی و یادگیری عمیق

دکتر صفابخش



دانشگاه صنعتی امیرکبیر
(پلی تکنیک تهران)
دانشکده مهندسی کامپیوتر

رضا آدینه پور ۴۰۲۱۳۱۰۵۵

تمرین چهارم
شبکه CNN

۲۰ اردیبهشت ۱۴۰۳



دانشکده مهندسی کامپیوتر

شبکه‌های عصبی و یادگیری عمیق

تمرین چهارم

رضا آدینه پور ۴۰۲۱۳۱۰۵۵

سوال اول - نظری

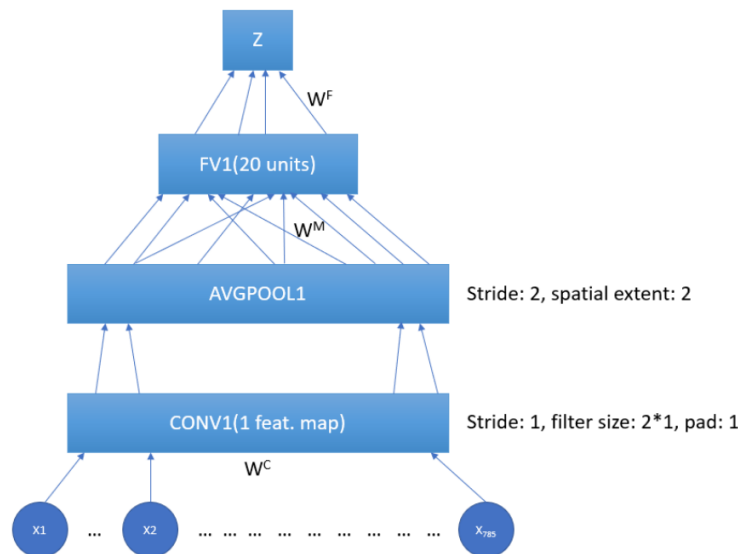
نحوه اشتراک گذاری پارمترها در لایه های کانولوشنی باعث ویژگی Equivariance نسبت به Translation می شود. این ویژگی را شرح دهید و کاربرد آن را توضیح دهید.

سوال دوم - نظری

شبکه‌های عمیق از عدم تفسیرپذیری رنج می‌برند. تلاش برای حل این مشکل، دو ایده Deconvolutional و Up-convolutional مطرح شده است. بررسی کنید و توضیح دهید هرکدام از دو روش، به چه صورت منجر به تفسیرپذیری می‌شوند؟

سوال سوم - نظری

معماری شبکه کانولوشنی زیر را در نظر بگیرید:



شکل ۱: شبکه کانولوشنی مورد بررسی در سوال سوم

- ابعاد ورودی 1×785 و خروجی شبکه 1×1
- لایه ورودی X با Zero-padding با طول ۱
- لایه کانولوشنی یک‌بعدی Conv1 با یک کرنل 1×2 و تابع فعال‌سازی ReLU
- لایه Average-polling (AVGPOOL1)
- لایه تمام متصل FC1 با تابع فعال‌سازی ReLU
- لایه خروجی Z که به لایه FC1 کاملاً متصل است و تابع فعال‌سازی Sigmoid

وزن لایه FC1 به Z را با W_i^F ، بایاس Z را با b^F ، وزن لایه AVGPOOL1 به FC1 را با W_{ij}^A ، بایاس FC1 را با b_i^M بردار W^C برابر $[W_1^C, W_2^C]$ و بایاس لایه کانولوشنی را با b^C نشان می‌دهیم. داده‌های مجموعه آموزش به صورت X^i و خروجی مورد انتظار به صورت Y^i است. همچنین خروجی‌های لایه‌های شبکه به ترتیب $c(X^i)$ ، $a(X^i)$ ، $f(X^i)$ ، $z(X^i)$ می‌نامیم. در این صورت، تابع هزینه به صورت زیر تعریف می‌شود:

$$\text{cost}(X, Y) = \sum_n \text{cost}(X^{(n)}, Y^{(n)}) = \sum_n (-Y^{(n)} \log(z(X^{(n)})) - (1 - Y^{(n)}) \log(1 - z(X^{(n)})))$$

باتوجه به مفروضات بالا، به پرسش‌های زیر پاسخ دهید:

۱. تعداد پارامترهای شبکه بالا را با ذکر جزئیات محاسبه کنید.
۲. برای فقط یک نمونه آموزشی، مقدار $\frac{\partial \text{Cost}}{\partial W_{ji}^A}$ و $\frac{\partial \text{Cost}}{\partial W_1^C}$ را با جزئیات محاسبه کنید.

سوال چهارم - نظری

کانولوشن متسع^۱ روشی برای افزایش میدان پذیرش (Receptive field) شبکه‌های کانولوشنی است که به صورت زیر تعریف می‌شود: (دقت شود خروجی تنها برای اندیس‌هایی از کرنل و تصویر همپوشانی کامل دارند، محاسبه می‌شود)

$$(k * I)(i, j) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} K(m, n) I(i + D_m, j + D_n)$$

۱. در یک شبکه کانولوشنی با یک لایه کانولوشن $K \times K$ با طول گام یک، عرض میدان پذیرش را بدست آورید.
۲. برای ورودی $I \in \mathbf{R}^{M \times N}$ و کرنل $K \in \mathbf{R}^{F \times F}$ ، نشان دهید خروجی عملگر متسع دارای ابعاد $(M - DF + D) \times (N - DF + D)$ است. متغیر D به معنی Dilation است.
۳. نشان دهید کانولوشن متسع معادل کانولوشن با کرنل متسع شده $K' = K \otimes A$ است. ماتریس A را مشخص کنید. (عملگر \otimes به معنی Kronecker product است.)

^۱ Dilated convolution

سوال پنجم - عملی

شبکه‌های کانولوشنی با توجه به توانایی آن‌ها در استخراج و یادگیری خودکار ویژگی‌ها، مقاومت نسبت به تغییرات و کارایی آن‌ها در مقابل پیچیدگی‌های وظیفه‌ی بازشناسی چهره، یک عنصر اساسی در اکثر اسن سیستم‌ها هستند. در این تمرین قصد داریم که با استفاده از شبکه‌های عصبی کانولوشنی به تحلیل احساسات چهره^۲ و طبقه‌بندی آن‌ها از روی تصویر بپردازیم. مجموعه داده‌ی این تمرین شامل ۱۲۰۰ تصویر نمونه‌گیری شده از هر کلاس مجموعه [AffectNet](#) می‌باشد. مجموعه داده AffectNet شامل ۴۵۰ هزار تصویر چهره با ۸ حالت مختلف می‌باشد که شکل ۲ نمونه‌هایی از آن را نشان می‌دهد.



شکل ۲: نمونه‌هایی از مجموعه داده AffectNet

^۲Facial expression recognition