

سوال ۱: Basic Concepts (۲۵ نمره)

۱. (۶ نمره) عملگرهای Convolution و Cross-Correlation گسسته دو بعدی، بین تصویر (I) و کرنل (K) به صورت زیر تعریف می‌شوند:

$$(K * I)(i, j) = \sum_{m, n} K(m, n) I(i - m, j - n) \quad (\text{Convolution})$$

$$(K \star I)(i, j) = \sum_{m, n} K(m, n) I(i + m, j + n) \quad (\text{Cross - Correlation})$$

الف) خاصیت جابه‌جایی (Commutativity) را در این دو عملگر بررسی کنید.
ب) در شبکه‌هایی که با استفاده از هر یک از این عملگرها پیاده‌سازی شده اند تفاوت وزن‌های یادگیری شده را بررسی کنید. آیا در عملکرد کلی شبکه تفاوتی ایجاد می‌شود؟

۲. (۵ نمره) یک تابع Pooling خروجی یک لایه شبکه کانولوشنی در یک مکان مشخص را با آماره‌ای از واحد (unit) های اطراف آن (Max, Average, ...) جایگزین می‌کند.

الف) با یک مثال نشان دهید Max-Pooling خروجی لایه کانولوشن را نسبت به جابه‌جایی مستقل (Invariant) می‌کند. (با جابه‌جایی جزئی مقادیر ورودی، خروجی لایه تغییر زیادی نمی‌کند)
ب) در چه کاربردهایی استفاده از Pooling ممکن است نتیجه مطلوب را در پی نداشته باشد؟

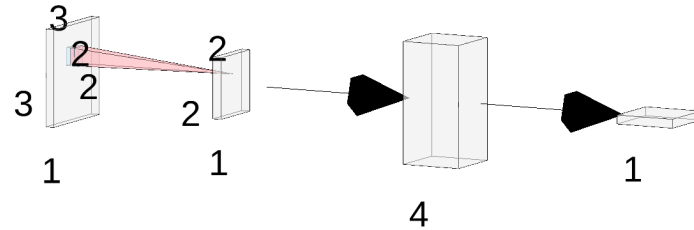
۳. (۶ نمره) در بیشتر کاربردها تابع مدل شده توسط شبکه عصبی یک توزیع پارامتریک $p(y|x; \theta)$ را مدل می‌کند (مسائل دسته‌بندی). نشان دهید تابع خطای Cross-Entropy که به صورت زیر تعریف می‌شود حاصل تخمین Maximum Likelihood پارامترهای این مدل است. N تعداد نمونه‌ها در مجموعه آموزش و K تعداد کلاس‌ها در مسئله دسته‌بندی و θ پارامترهای مسئله را مشخص می‌کنند. $y^{(i)}$ را به صورت one-hot در نظر بگیرید.

$$D = \{(x^{(i)}, y^{(i)})\}_{i=1}^N \sim^{iid} p_{x,y}$$

$$J(\theta) = - \sum_{i=1}^N \sum_{k=1}^K y_k^{(i)} \log p(y_k^{(i)} | x^{(i)}; \theta)$$

۴. (۸ نمره) الف) مشکلات اصلی تابع فعال‌ساز Sigmoid که باعث کند شدن فرایند یادگیری می‌شوند را توضیح دهید.
ب) دلیل برتری تابع خطای Cross-Entropy نسبت به Mean-Squared Error در مواقعی که شبکه $p(y|x; \theta)$ را مدل می‌کند توضیح دهید.^۱ (راهنمایی: بررسی کنید وجود لگاریتم در رابطه Cross-Entropy چه کمکی به فرایند آموزش می‌کند).

^۱ برای آشنایی بیشتر با تابع خطای Cross-Entropy به بخش 6.3 کتاب Deep Learning مراجعه کنید



۱. (۱۴ نمره) یک شبکه کانولوشنی با ورودی 3×3 ، یک لایه کانولوشن با یک فیلتر 2×2 و تابع فعالساز Relu و یک لایه تمام متصل با تابع فعالساز Sigmoid که در شکل آمده است را در نظر بگیرید.

الف) شبکه MLP معادل این شبکه را رسم کنید. این شبکه چه تفاوتی با شبکه های تمام متصل دارد.

ب) تعداد پارامترهای این شبکه را با شبکه تمام متصل (Fully-Connected) با ساختار یکسان مقایسه کنید.

پ) از خصوصیات شبکه های کانولوشنی Sparse Interactions و Shared Weights است. نشان دهید این خصوصیات حاصل یک فرض پیشین (Prior) قوی روی پارامترهای مدل است. این فرض پیشین در چه مسائلی درست است.^۲

ت) با فرض تابع خطای Cross-Entropy مشتق خطا نسبت به یکی از وزن های فیلتر کانولوشن را به دست آورید.

۲. (۶ نمره) الف) در صورت مقدار دهی اولیه تمام وزن ها در یک لایه شبکه کانولوشنی با یک مقدار ثابت یکسان چه مشکلی در فرایند آموزش ایجاد می شود. (راهنمایی: فرایند forward propagation و بروزرسانی وزن ها را در این حالت بررسی کنید)

ب) در یک شبکه کانولوشنی با تابع فعال سازی Sigmoid تمامی وزن ها با یک متغیر تصادفی نرمال با انحراف معیار ۱۰ مقدار دهی شده اند. ویژگی Vanishing Gradient چه مشکلی را در آموزش شبکه ایجاد میکند؟

^۲ برای آشنایی بیشتر با ویژگی های شبکه های کانولوشنی بخش 9.2 کتاب Deep Learning را مطالعه کنید