



مقدمه‌ای بر یادگیری ماشین

پاییز ۱۴۰۱

اساتید: علی شریفی، بهروز آذرخلیلی

دانشگاه صنعتی شریف

دانشکده‌ی مهندسی کامپیوتر

تاریخ برگزاری: -

فصل ۶

پاسخ کوییز دوم

سوالات (۱۰۰ نمره)

۱. (۲۰) به سوالات زیر پاسخ کوتاه دهید.

- دو تاثیر استفاده از Padding را نام ببرید.
- چرا ماژول Inception از کانولوشن یک در یک استفاده می کند؟
- تابع خطا Variational Autoencoder چه هدفی را علاوه بر تابع خطا Autoencoder دنبال می کند؟
- اعمال فیلتر $\begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$ بر روی یک عکس Grayscale در تشخیص چه چیزی کمک می کند؟

حل.

- حفظ ابعاد و تاثیر دادن بیشتر پیکسل های کناری در محاسبات.
- کم کردن محاسبات - برای دیدن مثال به اسلایدها مراجعه شود.
- در تابع خطا Variational Autoencoder علاوه بر ترم Reconstruction ، ترم Regularization نیز وجود دارد که توزیع Latent Space را به نرمال استاندارد نزدیک می کند.
- تشخیص لبه های افقی.

۲. (۱۵) درستی و نادرستی گزاره های زیر را با ذکر دلیل مشخص کنید.

- مدل Autoencoder جزو روش های Supervised است.
- لایه Pooling به دلیل نداشتن وزنی برای یادگیری تاثیری در Backpropagation ندارد.
- از ساختار Autoencoder می توان برای حذف کردن نویز ورودی استفاده کرد.

حل.

- غلط - Unsupervised و عملا از ورودی خود به عنوان Supervision استفاده می کند.
- غلط - زیرا همچنان نیاز به محاسبه مشتق نسبت به ورودی را داریم.
- درست - به اسلایدها مراجعه شود.

۳. (۶۰) به سوالات زیر پاسخ کامل دهید.

- ایده مطرح شده در مدل ResNet به دنبال حل چه مشکلی بوده است و چطور به آن پرداخته است؟

- توضیح دهید Autoencoder و PCA از نظر کاهش ابعاد چه تفاوتی با هم دارند و مزایای استفاده از Autoencoder نسبت به PCA چیست؟
- یک عکس RGB با ابعاد 300×300 را به عنوان ورودی در نظر بگیرید. تعداد پارامترها در یک لایه Dense با ۱۰۰ نورون (بدون احتساب bias) و یک لایه Convolution با ۱۰۰ فیلتر 5×5 (با احتساب bias) را بدست آورید.
- در کدنویسی وقتی از Padding در حالت Same در یک لایه کانولوشن با ۳۲ فیلتر 7×7 با Stride یک برای یک ورودی $3 \times 63 \times 63$ استفاده می کنیم مقدار Padding چه مقداری در نظر گرفته می شود؟

حل.

- مدل ResNet به دنبال حل مشکل بهینه سازی مدل های بسیار عمیق بوده است و بدین منظور از Skip Connection استفاده کرده است.
- Autoencoder ها انعطاف پذیری بیشتری دارند و با استفاده از توابع فعالساز غیرخطی که در آنها استفاده می شود می توان Transformation های غیرخطی انجام داد و ویژگی هایی که رابطه غیرخطی با ویژگی های اولیه دارند را به دست آورد (ویژگی هایی که در PCA به دست می آیند یک ترکیب خطی از ویژگی های اولیه هستند)، در روش های کاهش ابعاد مثل PCA ویژگی ها کاملاً Uncorrelated هستند ولی در Autoencoder ویژگی هایی که به دست می آیند می توانند با هم Correlation داشته باشند.
- تعداد پارامترها در قسمت اول برابر است با

$$300^2 \times 3 \times 100 = 27M$$

و در قسمت دوم برابر است با:

$$100 \times 5^2 \times 3 + 100 = 7600$$

- چون Stride یک است طبق رابطه ای در اسلاید ها داریم:

$$P = \frac{F-1}{2} = \frac{7-1}{2} = 3$$

۴. (۵) به سوال چهار گزینه ای زیر پاسخ دهید.

- کدام گزینه محاسبات انجام شده در بلوک ResNet را به درستی نشان می دهد؟ (منظور از g همان Activation Function است)

- ☐ $a^{[l+2]} = g(W^{[l+2]}g(W^{[l+1]}a^{[l]} + b^{[l+1]}) + b^{[l+2]} + a^{[l]} + a^{[l+1]}$
- ☐ $a^{[l+2]} = g(W^{[l+2]}g(W^{[l+1]}a^{[l]} + b^{[l+1]}) + b^{[l+2]})$
- ☐ $a^{[l+2]} = g(W^{[l+2]}g(W^{[l+1]}a^{[l]} + b^{[l+1]}) + b^{[l+2]}) + a^{[l]}$
- ☐ $a^{[l+2]} = g(W^{[l+2]}g(W^{[l+1]}a^{[l]} + b^{[l+1]}) + b^{[l+2]} + a^{[l]})$

حل. گزینه ۴ - به نوت بوک درس مراجعه شود.