فردین آیار

9)

ابتدا میانگین و انحراف معیار همه فیچرمپ ها محاسبه می­شود. سپس برای داشتن توزیع نرمال کل فیچرمپ ها منهای میانگین و تقسیم بر انحراف معیار میشوند(به طور دقیق تر تقسیم بر جذر واریانس به علاوه یک مقدار ابسیلون)

سپس کلیه فیچرمپ ها که دارای توزیع نرمال با میانگین صفر و واریانس یک شده اند با استفاده از دو پارامتر a و b به صورت ax+b به توزیع جدید منتقل می­شوند. (پارامتر های a و b در شبکه یادگرفته می­شوند)

نرمال سازی به این دلیل انجام میشود که با تغییر وزن های لایه های قبل ممکن است ورودی های لایه فعلی در هر مرحله دارای توزیع متفاوتی شود و در نتیجه یادگیری سخت تر میشود(=Internal covariate shift باعث سخت تر شدن تعیین ضریب یادگیری و پارامترهای اولیه نیز میشود)

منظور از دسته ای این است که کلیه فیچرمپ ها با هم نرمال میشوند و در واقع از میانگین و واریانس کل آن ها برای نرمال سازی استفاده می شود.

چون کلیه فیچر مپ ها با هم در لایه بعدی پردازش می­شوند این کار باید به صورت دسته ای انجام شود، در واقع هر درایه از فیچرمپ ها به صورت یک متغیر تصادفی است که برای نرمال سازی باید چندین مقدار از آن را داشته باشیم. یعنی باید چند فیچر مپ با هم نرمال شوند.