

به طور کلی هدف از این کار نرمالسازی وزن ها در سطح لایه و بین لایه ها است

و به این صورت عمل می کنند که batch یا یک بچه از داده ها را

به صورت زیر تبدیل می کنند

$$\mu_B = \frac{1}{n} \sum x_i$$

$$\sigma_B^2 = \frac{1}{n} \sum (x_i - \mu_B)^2$$

$$\hat{x}_i = \frac{x_i - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}}$$

$$y_i = \gamma \hat{x}_i + \beta = BN(x_i)$$

میانگین توزیع جدید  
انحراف معیار توزیع جدید

معنی در واقع در این جا به کمک ضرایب  $\gamma$  یا میانگین و توزیع  $\beta$  را داریم.

در شبکه های عمیق برای حل مشکل internal covariance shift از این روش استفاده

می کنند چون توزیع ورودی هر لایه دائم عوض می شود و تغییرات وزن ها زیاد است و آموزش را مختل می کند و همگرای سخت می شود به همین دلیل برای

تسریع روند آموزش BN انجام می شود، علت استفاده به صورت batch

این است که میانگین و واریانس به میانگین و واریانس توزیع کلی نزدیک باشد