

Batch Normalization

目次

1 順伝播

1

1 順伝播

学習にバッチサイズ m の入力を与えられるときの順伝播を考える。ネットワークの i 番目のニューロンへの λ 個目の入力を $x_{i,\lambda}^I$ 、同様に Batch Normalization layer への入力を $x_{i,\lambda}$ 、ネットワークの出力を $x_{i,\lambda}^O$ とする。順伝播の計算は

$$\begin{aligned}x_{i,\lambda} &= f_{i,\lambda}(x^I, \theta^I) \\ y_{i,\lambda} &= BN_{i,\lambda,\beta_i,\gamma_i}(x) \\ x_{i,\lambda}^O &= g_{i,\lambda}(o, \theta^O)\end{aligned}$$

として計算されるとする。ここで θ^I, θ^O はそれぞれ Batch Normalization より前と後の層のパラメータとした。また、Normalization の計算 $BN_{i,\lambda,\beta_i,\gamma_i}(x)$ は

$$\begin{aligned}\mu_i &= \frac{1}{m} \sum_{\lambda=1}^m x_{i,\lambda} \\ \sigma_i^2 &= \frac{1}{m} \sum_{\lambda=1}^m (x_{i,\lambda} - \mu_i)^2 \\ \hat{x}_{i,\lambda} &= \frac{x_{i,\lambda} - \mu_i}{\sqrt{\sigma_i^2 - \epsilon}} \\ y_{i,\lambda} &= \gamma_i x_{i,\lambda} + \beta_i\end{aligned}$$

として計算される。

また、学習の際はミニバッチによって平均と分散が計算されるが、実際の識別の場合は学習の最後のエポックで用いた平均と分散を用いる等して計算される。

参考文献

- [1] 斎藤康毅, “ゼロから作る Deep Learning”, オライリージャパン, (2016).