Normalization
Alex Tong
Advanced Topics in Data Mining and Machine Learning
4 April 2018

Normalization of the input data has always been an important step in the practical application of neural networks and learning models in general. It is unclear exactly what normalization entails, but the idea is to put inputs in the same domain, either uniformly over some range or normally with mean zero and unit variance.

In 2015, Sergey Ioffe and Christian Szegedy published a paper titled **Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift**. Since then, batch normalization has joined convolutional layers as defacto standard in networks on image related tasks. By normalizing inputs to internal layers of the network in addition to input data, batch normalization significantly improves training time and test error on image classification tasks.

Along with Batch Normalization, there has been recent work in other types of normalization such as Layer Normalization, Weight Normalization, Instance Normalization, and most recently, **Group Normalization**. In recent work out of Facebook, Yuxin Wu, and Kaiming He present group normalization as an alternative to batch normalization that normalizes over groups of channels instead of over batches. They claim that their method of normalization outperforms batch normalization in small batch size applications and performs comparably on other tasks.

Instead of directly adding a normalization layer into the network, another line of investigation is normalization through modifying the activation function. Work in **Self-Normalizing Neural Networks** out of Johannes Kepler University Linz in Austria examines this possibility. Through a modified activation unit, which they name "scaled exponential linear units" (SELU), they prove self-normalizing properties of internal layer weights. This allows them to construct deeper networks without vanishing/exploding gradients.

Readings:

[1] Ioffe, Sergey, and Christian Szegedy. "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift," 2015. ICML. https://arxiv.org/abs/1502.03167.

[2] Klambauer, Günter, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. "Self-Normalizing Neural Networks," 2017. NIPS. https://arxiv.org/abs/1706.02515v5.

[3] Wu, Yuxin, and Kaiming He. "Group Normalization," 2018. Tech Report. http://arxiv.org/abs/1803.08494.

Spotlight Question:
How applicable is normalization to domains outside of image recognition?