Gaussian Noise Passes Through Neural Network

Xiaocan Li
Institute of Automation, Chinese Academy of Sciences
No.95 Zhongguancun East Road
Haidian District, Beijing
lixiaocan2017@ia.ac.cn

Abstract

We study how gaussian noise behaves when it passes through different layers of neural network. We study the most widely used activation functions:Convolution(Linear activation), ReLU activation, Leaky ReLU activation, Tanh activation, Sigmoid activation, Max-pooling activation.

1. Introduction

The last 6 years have seen the rapid development of neural network[1][3][4] in machine learning community. As a common framework in deep learning theory, neural network has paved its way to object recognition and classification, as well as other applications. In [2] convolution and max-pooling operation on Gaussian noise[5], the most common noise in real world, are discussed. In the following section, more operations are involved.

2. Gaussian Noise and Widely Used Activations

2.1. Gaussian Noise and Convolution

Assume $X \sim N(0, \sigma^2)$, i.i.d, convolution is the weighted sum of each pixel value, $Y = \Sigma_{i=1}^N \omega_i x_i$. According to the property of Gaussian distribution, the PDF of Y is still Gaussian, its expected value and deviation:

$$E[Y] = \Sigma_{i=1}^N \omega_i E[x_i] = 0$$
$$D[Y] = \sigma^2 \Sigma_{i=1}^N \omega_i^2$$

Therefore, the PDF of Y:

$$f_Y(t) = \frac{1}{\sqrt{2\pi\Sigma_{i=1}^N \omega_i^2}\sigma} exp(-\frac{t^2}{2\sigma^2 \Sigma_{i=1}^N \omega_i^2})$$

2.2. Gaussian Noise and ReLU Activation

Assume $X \sim N(0, \sigma^2)$, i.i.d,
$Y = ReLU(x) = max(0, x)$.

Then the CDF of Y:

$$\begin{aligned} F_Y(t) &= P(Y \leq t) \\ &= P(max(0, x) \leq t) \\ &= P(x \leq t, 0 \leq t) \\ &= P(x \leq t), (t \geq 0) \\ &= F_x(t), (t \geq 0) \end{aligned}$$

Therefore, the PDF of Y:

$$\begin{aligned} f_Y(t) &= \frac{dF_Y(t)}{dt} \\ &= f_x(t) \\ &= \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{t^2}{2\sigma^2}), (t \geq 0) \end{aligned}$$

2.3. Gaussian Noise and Leaky ReLU Activation

Assume $X \sim N(0, \sigma^2)$, i.i.d,

$$Y = LeakyReLU(x) = \begin{cases} x & x \geq 0 \\ \alpha x & x < 0 \end{cases}$$

Then the CDF of Y:

$$\begin{aligned} F_Y(t) &= P(Y \leq t) \\ &= P(x \leq t, x \geq 0) + P(\alpha x \leq t, x < 0) \\ &= \begin{cases} P(0 \leq x \leq t) + P(x < 0) & t \geq 0 \\ P(x \leq \frac{t}{\alpha}) & t < 0 \end{cases} \\ &= \begin{cases} F_x(t) & t \geq 0 \\ F_x(\frac{t}{\alpha}) & t < 0 \end{cases} \end{aligned}$$

Therefore, the PDF of Y:

$$
\begin{aligned}
f_Y(t) &= \frac{dF_Y(t)}{dt} \\
&= \begin{cases} f_x(t) & t \geq 0 \\ \frac{1}{\alpha} f_x(\frac{t}{\alpha}) & t < 0 \end{cases} \\
&= \begin{cases} \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{t^2}{2\sigma^2}) & t \geq 0 \\ \frac{1}{\sqrt{2\pi}\alpha\sigma} exp(-\frac{t^2}{2\alpha^2\sigma^2}) & t < 0 \end{cases}
\end{aligned}
$$

We can see that when $\alpha \rightarrow 0$, Leaky ReLU becomes zero ReLU discussed in previous section, i.e. Leaky ReLU is a generalized version of zero ReLU.

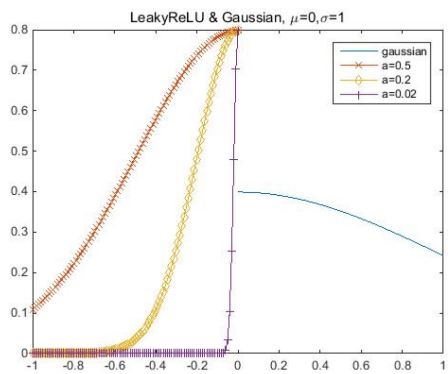

Figure 1. LeakyReLU and Gaussian

2.4. Gaussian Noise and Tanh Activation

Assume $X \sim N(0, \sigma^2)$, i.i.d,
$Y = tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$.
Then the CDF of Y:

$$
\begin{aligned}
F_Y(t) &= P(Y \leq t) \\
&= P(\frac{e^x - e^{-x}}{e^x + e^{-x}} \leq t) \\
&= P(x \leq \frac{1}{2} ln \frac{1+t}{1-t}), (|t| < 1)
\end{aligned}
$$

Therefore, the PDF of Y:

$$
\begin{aligned}
f_Y(t) &= \frac{dF_Y(t)}{dt} \\
&= \frac{1}{1-t^2} f_x(\frac{1}{2} ln \frac{1+t}{1-t}) \\
&= \frac{1}{\sqrt{2\pi}\sigma(1-t^2)} exp(-\frac{(ln\frac{1+t}{1-t})^2}{8\sigma^2}), (|t| < 1)
\end{aligned}
$$

Expected value of Y:$E[Y] = 0$ due to $tanh(x)$ is an odd function.

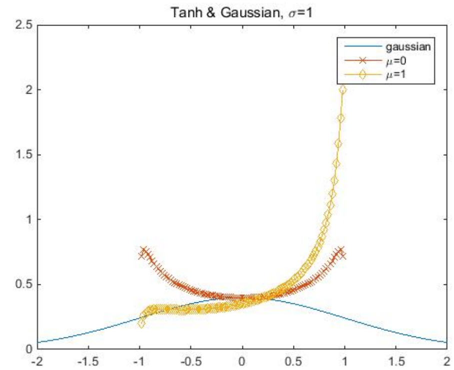

Figure 2. Tanh and Gaussian, tune μ

When we fix the standard deviation $\sigma = 1$ and tuning the parameter μ, we can see that the distribution will be symmetric about vertical axis $x = 0$ if $\mu = 0$; However, if $\mu \neq 0$, the distribution will be asymmetric, inclined to the left.

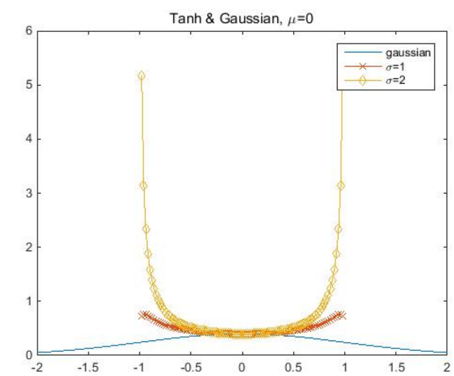

Figure 3. Tanh and Gaussian, tune σ

When we fix mean $\mu = 0$ and tuning the parameter σ, the distribution will always be symmetric about the vertical axis $x = 0$. Also, the greater σ is, the more flattened the distribution will be.

2.5. Gaussian Noise and Sigmoid Activation

Assume $X \sim N(0, \sigma^2)$, i.i.d,
$Y = sigmoid(x) = \frac{1}{1+e^{-x}}$.
Then the CDF of Y:

$$
\begin{aligned}
F_Y(t) &= P(Y \leq t) \\
&= P(\frac{1}{1+e^{-x}} \leq t) \\
&= P(x \leq ln\frac{t}{1-t}) \\
&= F_x(ln\frac{t}{1-t}), (0 < t < 1)
\end{aligned}
$$

Therefore, the PDF of Y:

$$
\begin{aligned}
f_Y(t) &= \frac{dF_Y(t)}{dt} \\
&= \frac{1}{t(1-t)} f_x(ln\frac{t}{1-t}) \\
&= \frac{1}{\sqrt{2\pi}\sigma t(1-t)} exp(-\frac{(ln\frac{t}{1-t})^2}{2\sigma^2}), (0 < t < 1)
\end{aligned}
$$

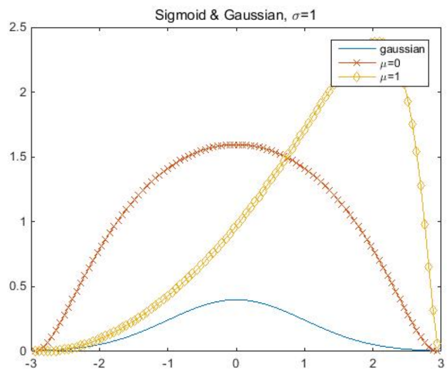

Figure 4. Sigmoid and Gaussian, tune μ

When we fix the standard deviation $\sigma = 1$ and tuning the parameter μ, the distribution will be pushed to the right, the mean value is larger as well. It is symmetric about vertical axis $x = 0$ when $\mu = 0$.

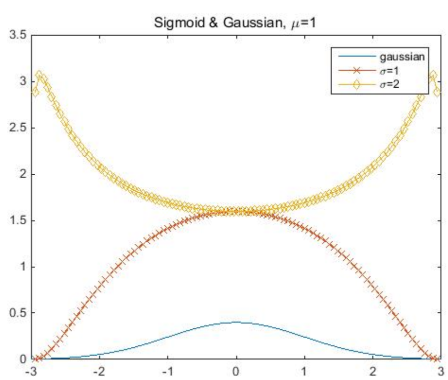

Figure 5. Sigmoid and Gaussian, tune μ

In Fig 5, when we fix the mean $\mu = 0$ and tuning the parameter σ, the distribution is always symmetric about the vertical axis $x = 0$ regardless of σ. But interestingly, the distribution of $\sigma = 1$ and $\sigma = 2$ looks flipped upside down.

2.6. Gaussian Noise and Max-pooling

Assume $X \sim N(0, \sigma^2)$, i.i.d, $Y = max(x_1, x_2, ..., x_N)$. Then the CDF of Y:

$$
\begin{aligned}
F_Y(t) &= P(Y \leq t) \\
&= P(x_1 \leq t, x_2 \leq t, ..., x_N \leq t) \\
&= F_{x_1}(t)F_{x_2}(t)...F_{x_N}(t) \\
&= F_x^N(t)
\end{aligned}
$$

Therefore, the PDF of Y:

$$
\begin{aligned}
f_Y(t) &= \frac{dF_Y(t)}{dt} \\
&= NF_x^{N-1}(t)f_x(t)
\end{aligned}
$$

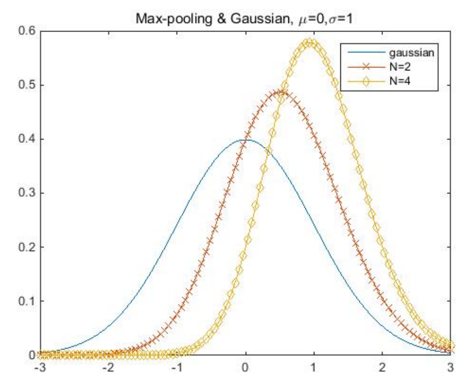

Figure 6. Maxpooling and Gaussian, tune N

Fix $\sigma = 1$ and $\mu = 0$, tune N. When N increases, the mean value shifts to the right and the peak value goes up.

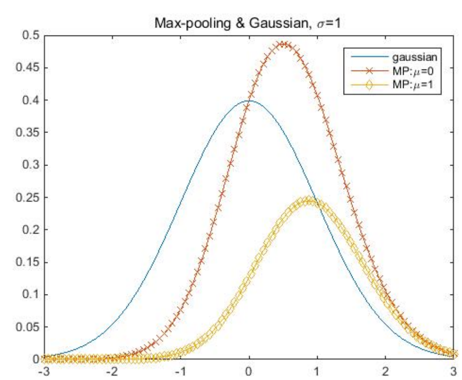

Figure 7. Maxpooling and Gaussian, tune μ

Fix $\sigma = 1$ and $N = 2$, tune μ. When μ increases, the mean value shifts to the right and the peak value goes down.

Fix $\mu = 0$ and $N = 2$, tune σ. When σ increases, the mean value shifts to the left and the peak value goes down.

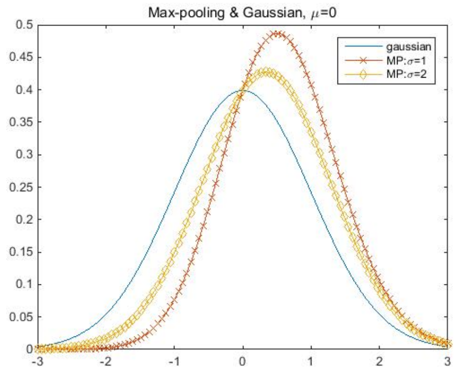

Figure 8. Maxpooling and Gaussian, tune σ

3. Gaussian Noise Passes Through Multiple Layers

We have discussed Gaussian noise passing through single layer of several commonly used activations. In practice, the number of hidden layers is between 4 to 20 or more. Thus, the feature of Gaussian noise passing through multiple layers needs to be found.

The input layer is convolutionized(weighted sum) before it is passed to next activation function, Max-pooling. According to Section 2.1, it is still a Gaussian. After Max-pooling operation, the PDF of output follows what is derived in Section 2.7. Next, these distributions are convolutionized again, we obtained another Gaussian distribution thanks to the power of Central Limit Theorem. Then, it is the very Gaussian distribution passes through convolutional layer and Max-pooling layer, which is an identical manipulation aforementioned. After several times of convolution and Max-pooling, fully connected layer is encountered, and we believe the output will be a Gaussian for the sake of the ubiquitous Central Limit Theorem.

Finally we meet the last layer of neural network, Softmax activation layer. Unfortunately the analytical solution of distribution is untractable.

4. Conclusion

The PDF of Gaussian noise passing through the most widely used activation functions is derived. These activations are Convolution(Linear activation), ReLU activation, Leaky ReLU activation, Tanh activation, Sigmoid activation and Max-pooling activation. More work needs to be done, such as experiments on real CNN, adding Gaussian noise to image to see if it still works.

References

[1] G. H. A Krizhevsky, I Sutskever. Imagenet classification with deep convolutional neural networks. In *Neural Information Processing Systems*, 2012.

[2] E. C. J Jin, A Dundar. Robust convolutional neural networks under adversarial noise. In *ICLR*, 2016.

[3] A. Z. Karen Simonyan. Going deeper with convolutions. In *ICLR*, 2015.

[4] A. Z. Karen Simonyan. Very deep convolutional neural networks for large-scale image recognition. In *ICLR*, 2015.

[5] Wikipedia. Gaussian distribution. `https://en.wikipedia.org/wiki/Normal_distribution`.