Correlation and Convolution in neural networks derivation

Nasy

Jan 27, 2023

Outline

Two operators $*$ and $\star$

- $(f \star g)(\tau) := \int_{-\infty}^{\infty} f(t) g(t+\tau)\, dt$
- $(f * g)(\tau) := \int_{-\infty}^{\infty} f(t) g(\tau - t)\, dt$

And for discrete functions:

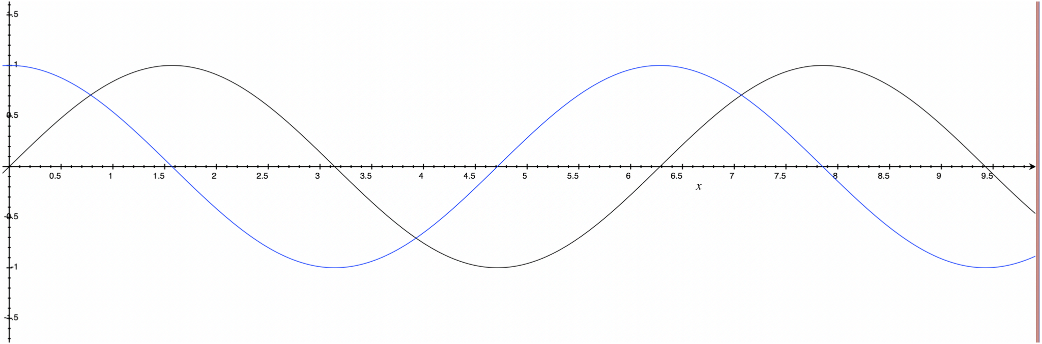
- $(f \star g)(\tau) := \sum_{-\infty}^{\infty} f(t) g(t+\tau)$
- $(f * g)(\tau) := \sum_{-\infty}^{\infty} f(t) g(\tau - t)$

Why?

Problem 1

Two signals f and g are given. They are from a same audio source, but there is a difference in the time when the recording was started. We want to align them.

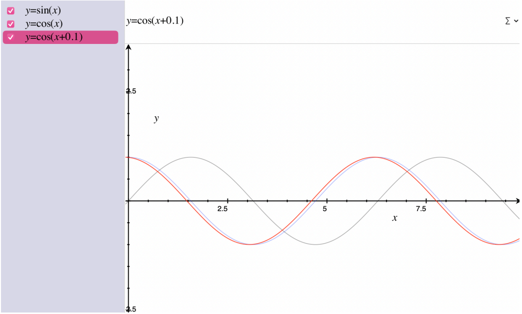
- $f(x) = \alpha \sin(x)$
- $g(x) = \beta \cos(x)$



Problem 1

Two signals f and g are given. They are from the same piece of audio, but there is a difference in the time when the recording was started. We want to align them.

- Same source, same crests and troughs.
- Keep f, shift g to align with f.
- $arg\max_{\Delta t}\sum_{-infty}^{infty} f(t)g(t+\Delta t)$



Problem 2

There is a function that describes the case where Mr. N is hit.

$$f(t) = 0, 0, 0, 1, 0, 0, 0, \ldots$$

And a function describing the change in pain after being hit.

$$g(t) = 3, 2, 1, 0$$

We want a function $h(t)$ to describe the pain of Mr. N.

Problem 2

There is a function that describes the case where Mr. N is hit.

$$f(t) = 0, 0, 0, 1, 0, 0, 0, \ldots$$

And a function describing the change in pain after being hit.

$$g(t) = 3, 2, 1; \;\; t \in \{0, 1, 2\}$$

$$g(t) = 0; \;\; t \in others$$

We want a function $h(\tau)$ to describe the pain of Mr. N.

- Pain = hit * pain
- At the first time hit, where $f(t) = f(3) = 1$, the pain is $g(0) = 3$.
- Next, the pain is $g(1) = 2$, and so on.

Obviously

$$h(\tau) = 0, 0, 0, 3, 2, 1, 0, 0, \ldots$$

What if $f(t) = 0, 0, 1, 1, 0, 0, \ldots$?

$$h(\tau) = 0, 0, 3, 5, 3, 1, 0, 0, \ldots$$

- $h(\tau) = \sum f(t) g(\tau - t)$

Forward (1D)

Correlation $\texttt{Input}(a) \star \texttt{kernel}(\texttt{filter})(w) = \texttt{output}(z)$

$$[0,1,2,3] \star [1,2] = [0 \times 1 + 1 \times 2, 1 \times 1 + 2 \times 2, 2 \times 1 + 3 \times 2] = [2,5,8]$$

$$[a_1, a_2, a_3, a_4] \star [w_1, w_2, w_3] = [a_1 \times w_1 + a_2 \times w_2 + a_3 \times w_3,\ a_2 \times w_1 + a_3 \times w_2 + a_4 \times w_3] = [z_1, z_2]$$

Backward (1D)

The δ_i is the gradient from the next layer at point i. And J is the loss function.

$$\frac{\partial J}{\partial a_i} = \sum_j \frac{\partial J}{\partial z_j}\frac{\partial z_j}{\partial a_i} = \sum_j \delta_j w_j$$

Thus,

$$\frac{\partial J}{\partial a_1} = \delta_1 w_1$$

$$\frac{\partial J}{\partial a_2} = \delta_1 w_2 + \delta_2 w_1$$

$$\frac{\partial J}{\partial a_3} = \delta_1 w_3 + \delta_2 w_2$$

$$\frac{\partial J}{\partial a_4} = \delta_2 w_3$$

$$[\delta_1 w_1, \delta_1 w_2 + \delta_2 w_1, \delta_1 w_3 + \delta_2 w_2, \delta_2 w_3]$$

Bcakward (1D)

$$\frac{\partial J}{\partial a_1} = 0w_3 + 0w_2 + \delta_1 w_1$$

$$\frac{\partial J}{\partial a_2} = 0w_3 + \delta_1 w_2 + \delta_2 w_1$$

$$\frac{\partial J}{\partial a_3} = \delta_1 w_3 + \delta_2 w_2 + 0w_1$$

$$\frac{\partial J}{\partial a_4} = \delta_2 w_3 + 0w_2 + 0w_1$$

$$[0, 0, \delta_1, \delta_2, 0, 0] * [w_1, w_2, w_3]$$

$$= [0, 0, \delta_1, \delta_2, 0, 0] \star [w_3, w_2, w_1] = [\delta_1 w_1, \delta_1 w_2 + \delta_2 w_1, \delta_1 w_3 + \delta_2 w_2, \delta_2 w_3]$$

2D correlation and convolution

Correlation $\texttt{Input} \star \texttt{kernel}(\texttt{filter}) = \texttt{output}$

$$\begin{bmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix} \star \begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 19 & 25 \\ 37 & 43 \end{bmatrix}$$

The shadow part:

$0 \times 0 + 1 \times 1 + 3 \times 2 + 4 \times 3 = 19$

Convolutions $\texttt{Input} * \texttt{kernel}(\texttt{filter}) = \texttt{Input} \star \mathit{flip}(\texttt{kernel}(\texttt{filter})) = \texttt{output}$

$$\begin{bmatrix} 0 & 1 & 2 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix} * \begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix} = \star \begin{bmatrix} 3 & 2 \\ 1 & 0 \end{bmatrix}$$

Forward (2D)

Correlation $\texttt{Input}(a) \star \texttt{kernel}(\texttt{filter})(w) = \texttt{output}(z)$

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \star \begin{bmatrix} w_{11} & w_{12} \\ w_{21} & w_{22} \end{bmatrix} = \begin{bmatrix} z_{11} & z_{12} \\ z_{21} & z_{22} \end{bmatrix} \tag{1}$$

$$z_{11} = a_{11}w_{11} + a_{12}w_{12} + a_{21}w_{21} + a_{22}w_{22}$$

$$z_{12} = a_{12}w_{11} + a_{13}w_{12} + a_{22}w_{21} + a_{23}w_{22}$$

$$z_{21} = a_{21}w_{11} + a_{22}w_{12} + a_{31}w_{21} + a_{32}w_{22}$$

$$z_{22} = a_{22}w_{11} + a_{23}w_{12} + a_{32}w_{21} + a_{33}w_{22}$$

Backward (2D)

$$\nabla a^{l-1} = \frac{\partial e}{\partial a^{l-1}} = \frac{\partial e}{\partial z^l}\frac{\partial z^l}{\partial a^{l-1}} = \delta^l \frac{\partial z^l}{\partial a^{l-1}} = \delta^l W$$

$$\nabla a_{11} = \delta_{11} w_{11}$$

$$\nabla a_{12} = \delta_{11} w_{12} + \delta_{12} w_{11}$$

$$\nabla a_{13} = \delta_{12} w_{12}$$

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & \delta_{11} & \delta_{12} & 0 \\ 0 & \delta_{21} & \delta_{22} & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \star \begin{bmatrix} w_{22} & w_{21} \\ w_{12} & w_{11} \end{bmatrix} = \begin{bmatrix} \nabla a_{11} & \nabla a_{12} & \nabla a_{13} \\ \nabla a_{21} & \nabla a_{22} & \nabla a_{23} \\ \nabla a_{31} & \nabla a_{32} & \nabla a_{33} \end{bmatrix} \tag{2}$$

Conv and Corr General Calculation

$$\texttt{Input}(a) \star \texttt{kernel}(\texttt{filter})(w) = \texttt{output}(z)$$

Basic method Direct calculation

$$\sum_{i=1}^{n} a_i w_i = z$$

It's an inner product! $< A, W >$. And the time complexity is $O(n^2) = O(N_A M_W)$.

Faster one FFT

$$y[n] = f[n] * g[n] \leftrightarrow Y[f] = F[f]\, G[f]$$

The time complexity is $O(n \log_2 n)$

Conclusion

- Correlation
- Convolution
- Forward and Backward
- Conv and Corr Calculations