Gaussian-Lorenzian Layer Background Removal

Ali Qajar
ali.qajar @ gmail.com

In the training of Multi-Layer Perceptrons (MLPs), weights of a layer are typically initialized with random noise and converge to near-optimal weights during training. However, MLPs are not efficient at learning transformations accurately. Instead, Residual Networks (ResNets) are preferred, where the network learns residuals (deviations from a background). Despite this, background information can still seep into the layer weights, reducing efficiency, increasing size, and prolonging training time.

To address this, we propose adding new conditions to the objective function to incentivize the network to learn only the residuals and separate them from the background. Inspired by scattering analysis techniques such as X-ray Photoelectron Spectroscopy (XPS), this approach aims to force a layer to pass more information through fewer neurons, thereby reducing background noise.

Background

Figure 1 shows an XPS spectrum of a surface with alkaline ions, where peaks represent direct photon-atom interactions and inelastic scattering tails indicate various interactions like electron-electron and electron-phonon interactions.

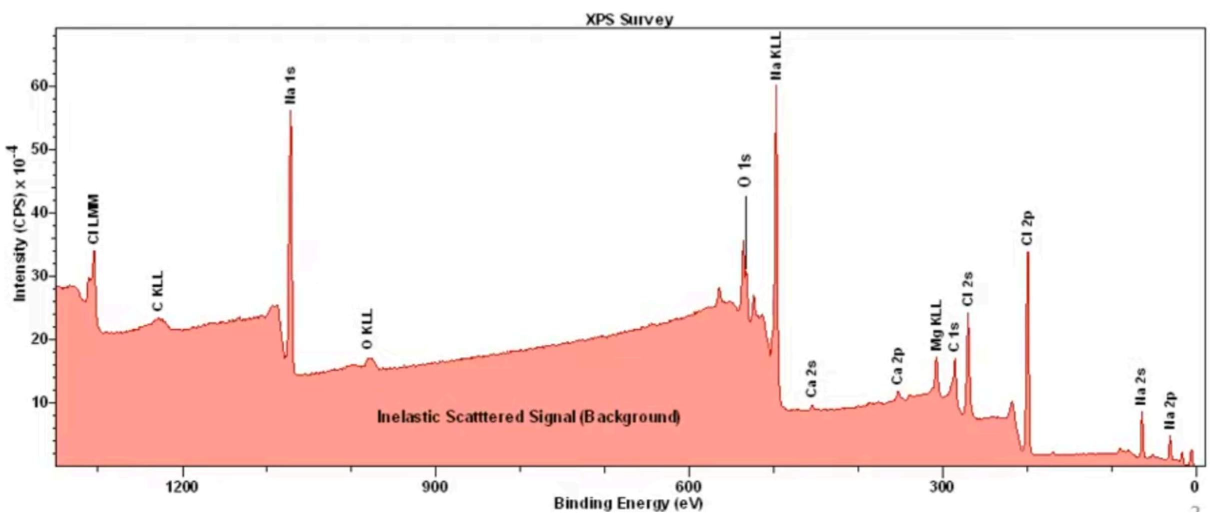


Figure 1: XPS of surface with alkaline ions. Peaks indicate the ions and Auger lines. Inelastic scattering shows up as a noisy background [1].

Figure 2 illustrates raw XPS data before and after background removal, necessary for accurate calculation of constituent components.

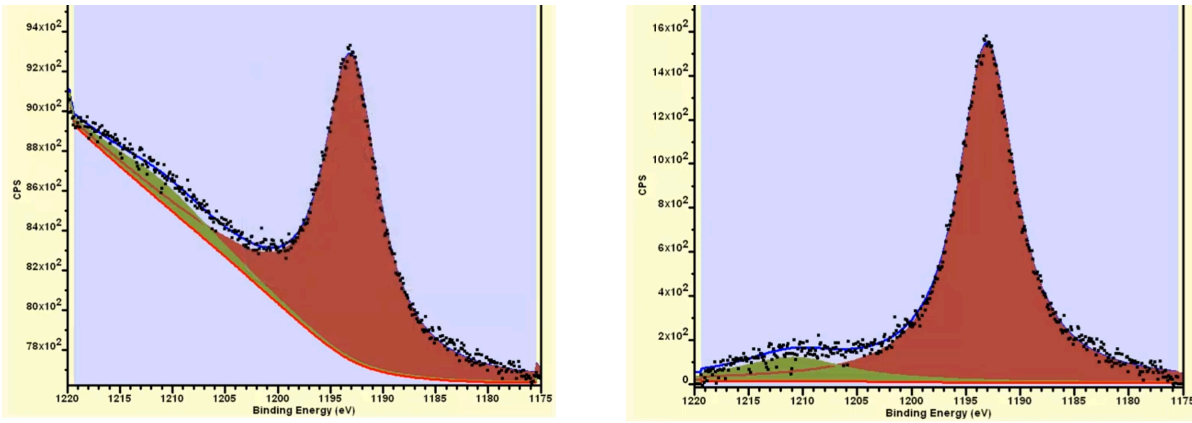


Figure 2: Left - raw data, Right - adjusted for background [2]

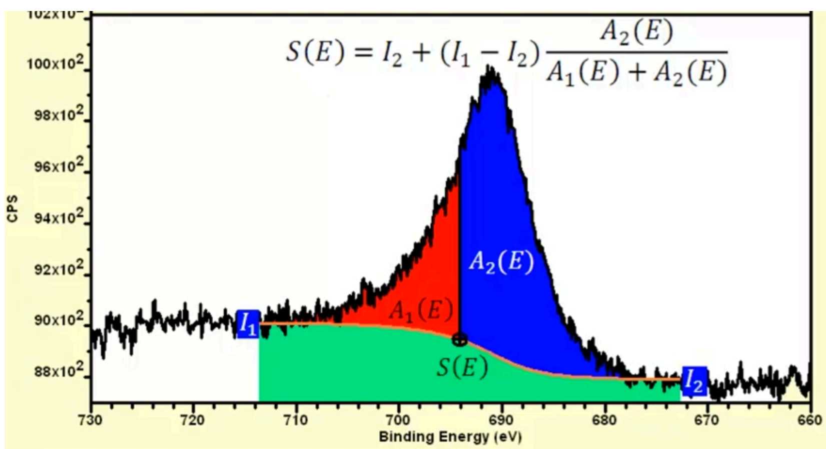


Figure 3: Background S(E) is computed. The area above the background is attributed to the constituent components.

Figure 3 depicts the computation of the Shirley background, which separates the inelastic scattering background from the core-level peaks.

Methodology

Background Removal in Neural Networks:

Inspired by the Shirley background removal method in XPS, we propose a similar approach for neural network layers. This involves computing the background and subtracting it from the total signal to isolate the residuals.

Formula for Shirley Background Calculation:

$$A_2(E) = \int_{E}^{E_2} J(x) - B_{sh}(x)dx$$

$$A_1(E) + A_2(E) = \int_{E_1}^{E_2} J(x) - B_{sh}(x)dx$$

Given a spectrum $J(E)$ and interval $[E_1, E_2]$, the background signal is defined such that $B_{sh}(E_1) = I_1$ and $B_{sh}(E_2) = I_2$, then the Shirley background is calculated by iteration using formula

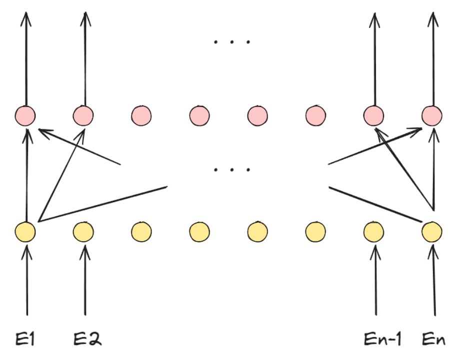
$$B_{sh_{n+1}}(E) = I_2 + \frac{(I_1 - I_2)}{\int_{E_1}^{E_2} J(x) - B_{sh_n}(x)dx} \int_{E}^{E_2} J(x) - B_{sh_n}(x)dx$$

If signal above inelastic scatter background is Lorenzian then without iteration the background equivalent to a Shirley background is given by

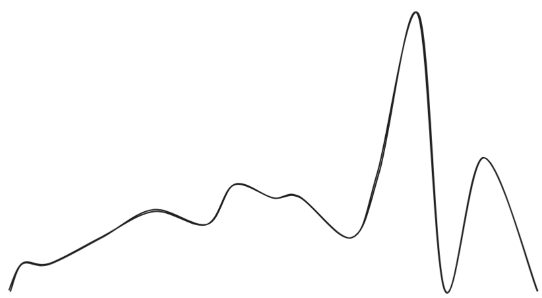
$$B_L(E) = I_2 + \frac{(I_1 - I_2)}{\int_{-\infty}^{\infty} 1/(1+x^2)dx} \int_{E}^{\infty} 1/(1 + x^2)dx = I_2 + \frac{(I_1 - I_2)}{\pi} (\pi/2 - tan^{-1}E)$$

Application to Neural Networks

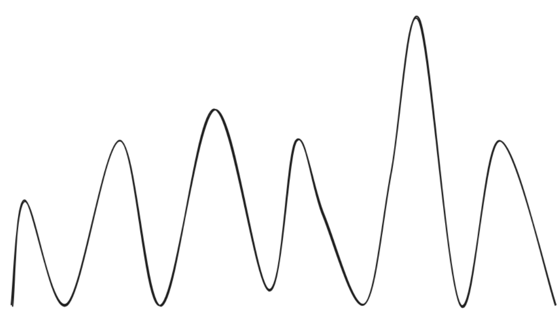
Figure 4 schematically represents the effect of this change. By detecting and removing the background, the layer can focus on learning the residual information, resulting in a more defined weight distribution.



(a)



(b)



(c)

Figure 4: (a) example layer in a ResNet. (b) Naive training (initiated with weight gaussian randomization). This model has no safeguard to prevent the layer from learning background noise. (c) Distribution of layer weights after applying Shirley background removal to the optimization objective. This model incentivises the layer to learn Residual features and increase signal to noise ratio.

Conclusion

Incorporating background removal techniques similar to those used in XPS into neural network training can enhance efficiency, reduce the size of networks, and accelerate training. This approach promotes a more effective learning of residuals, enabling the construction of more compact and efficient neural networks.

References

[1] CasaXPS Casa Software. https://www.youtube.com/@casaxpscasasoftware4605