

Chapter 11: Summary

Diedrick Darrell Darmadi - 1103223031

Bab 11 membahas **teknik melatih jaringan saraf dalam (deep neural networks)** dan tantangan yang muncul ketika jumlah layer meningkat. Bab ini dimulai dengan masalah **vanishing dan exploding gradients**, di mana gradien yang terlalu kecil atau besar menyebabkan training menjadi lambat atau tidak stabil.

Selanjutnya dijelaskan **fungsi aktivasi modern**, seperti ReLU, Leaky ReLU, ELU, dan SELU, yang membantu mengurangi masalah vanishing gradient. Grafik masing-masing fungsi aktivasi divisualisasikan untuk menunjukkan perbedaan perilaku terhadap input.

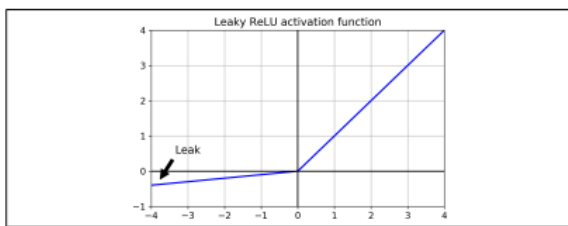


Figure 11-2. Leaky ReLU

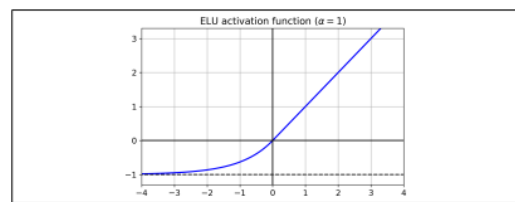


Figure 11-3. ELU activation function

Bab ini kemudian membahas **inialisasi bobot (weight initialization)** yang tepat, seperti Xavier/Glorot dan He initialization, yang penting untuk menjaga distribusi output layer tetap stabil di awal training.

Selanjutnya diperkenalkan **batch normalization**, teknik yang menormalkan input tiap batch untuk mempercepat training dan menstabilkan distribusi aktivasi. Berikut algoritmanya

Equation 11-3. Batch Normalization algorithm

$$\begin{aligned} 1. \quad \mu_B &= \frac{1}{m_B} \sum_{i=1}^{m_B} \mathbf{x}^{(i)} \\ 2. \quad \sigma_B^2 &= \frac{1}{m_B} \sum_{i=1}^{m_B} (\mathbf{x}^{(i)} - \mu_B)^2 \\ 3. \quad \hat{\mathbf{x}}^{(i)} &= \frac{\mathbf{x}^{(i)} - \mu_B}{\sqrt{\sigma_B^2 + \epsilon}} \\ 4. \quad \mathbf{z}^{(i)} &= \gamma \otimes \hat{\mathbf{x}}^{(i)} + \beta \end{aligned}$$

Bab ini juga membahas **gradient clipping**, **regularisasi (L1, L2, dropout)**, dan **early stopping** sebagai cara mengatasi overfitting dan menjaga stabilitas training.

Selanjutnya dijelaskan **optimizer modern** seperti Momentum, RMSProp, dan Adam, serta dampaknya terhadap kecepatan konvergensi.