

D14 : CNN分類器架構：Batch Normalization | Cupoy

 cupoy.com/event/cvdl/mission/1586226055085



Day 14

深度學習與電腦視覺
學習馬拉松

Cupoy 陪跑專家：楊哲寧

深度學習理論與實作



深度學習理論與實作

CNN原理：Batch Normalization

重要知識點

- 理解 Batch Normalization 原理

- Batch Normalization 用來解決什麼問題

BN (Batch Normalization)

Batch Normalization 是 2015 年 Google 研究員在論文《Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift》一文中提出的, 並同時將 BN 應用於 Inception-v2 的框架中。

BN算法

首先計算輸入 Batch 的平均值與標準差

Input: Values of x over a mini-batch: $\mathcal{B} = \{x_{1...m}\}$;
Parameters to be learned: γ, β
Output: $\{y_i = \text{BN}_{\gamma, \beta}(x_i)\}$

$$\mu_{\mathcal{B}} \leftarrow \frac{1}{m} \sum_{i=1}^m x_i \quad // \text{ mini-batch mean}$$
$$\sigma_{\mathcal{B}}^2 \leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu_{\mathcal{B}})^2 \quad // \text{ mini-batch variance}$$

$$\hat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} \quad // \text{ normalize}$$

$$y_i \leftarrow \gamma \hat{x}_i + \beta \equiv \text{BN}_{\gamma, \beta}(x_i) \quad // \text{ scale and shift}$$

透過學習 Gamma 與 Beta 做縮放與平移, Gamma 與 Beta 為 BN 層內唯二需要學習的參數

Input: Values of x over a mini-batch: $\mathcal{B} = \{x_1 \dots x_m\}$;
Parameters to be learned: γ, β

Output: $\{y_i = \text{BN}_{\gamma, \beta}(x_i)\}$

$$\mu_{\mathcal{B}} \leftarrow \frac{1}{m} \sum_{i=1}^m x_i \quad // \text{ mini-batch mean}$$

$$\sigma_{\mathcal{B}}^2 \leftarrow \frac{1}{m} \sum_{i=1}^m (x_i - \mu_{\mathcal{B}})^2 \quad // \text{ mini-batch variance}$$

$$\hat{x}_i \leftarrow \frac{x_i - \mu_{\mathcal{B}}}{\sqrt{\sigma_{\mathcal{B}}^2 + \epsilon}} \quad // \text{ normalize}$$

$$y_i \leftarrow \gamma \hat{x}_i + \beta \equiv \text{BN}_{\gamma, \beta}(x_i) \quad // \text{ scale and shift}$$

資料分佈

一般來說我們都是以 **Mini Batch** 的方式訓練資料，然而每一個 Batch 間的資料分佈可能不太相同，而輸入每一層神經元的資訊分布也都可能會改變，造成收斂上的困難。

透過 BN，將每一層輸入資料的分佈歸一化為平均值為 0，方差為 1，確保資料分佈的穩定性。

然而 Normalize 改變資料的分佈，可能會造成上一層學到的特徵消失，因此 BN 的最後一步透過學習 Beta、Gamma，去微調 Normalize 後資料的分佈。

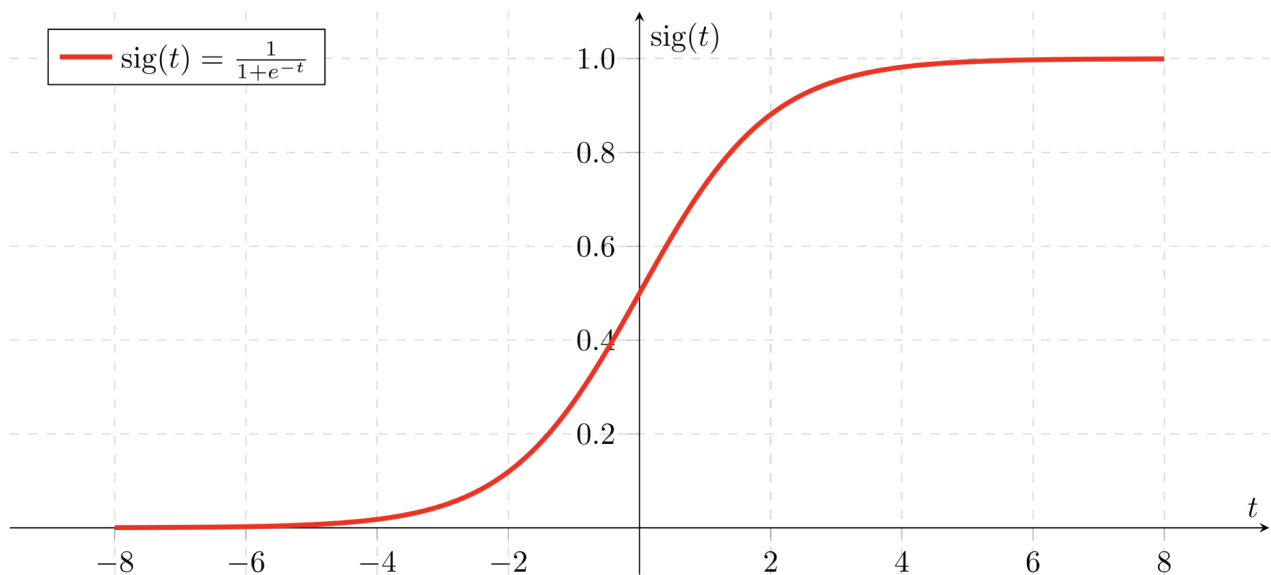
參考資料：

[Understanding the backward pass through Batch Normalization Layer](#)

[At the moment there is a wonderful course running at Stanford University, called CS231n - Convolutional Neural Networks for Visual Recognition, held by Andrej Karpathy, Justin Johnson and Fei-Fei Li. Fortunately all the course material is provided for free and all the lectures are recorded and uploaded on Youtube. This...](#)

kratzert.github.io

梯度消失



參考來源：Derivative of the Sigmoid function

Sigmoid會將數值較大與較小的值通通壓在一起，並且由於其導函數最大值為0.25，容易發生梯度消失的情形，透過BN，我們將資料分布歸一化，能有效降低梯度消失的可能性。

推薦延伸閱讀

解題時間

解題時間 Let's Crack It



Sample Code & 作業 開始解題