Khái niệm Epoch

Tìm hiểu khái niệm epochs và steps\_per\_epoch trong hàm fit\_generator của thư viện deeplearning Keras.

|  |
| --- |
| fit\_generator(generator, steps\_per\_epoch=**None**, epochs=1, verbose=1,  callbacks=**None**, validation\_data=**None**,  validation\_steps=**None**, validation\_freq=1,  class\_weight=**None**, max\_queue\_size=10, workers=1,  use\_multiprocessing=**False**, shuffle=**True**,  initial\_epoch=0) |

Là hàm huấn luyện model dựa trên data được tạo theo từng batch này qua batch khác.

Tham số generator trong hàm trên là đối tượng sinh ra các batch dữ liệu dùng để huấn luyện, giả sử 1 batch có 32 mẫu ta nói kích thước của batch là 32. Batch cuối cùng thường nhỏ hơn các batch trước đó vì lượng dữ liệu mẫu không chia hết cho kích thước của batch.

**Epoch:**

Giá trị truyền vào là một số nguyên.

Một Epoch là khi tất cả dữ liệu được duyệt qua một lần.

Khi dữ liệu quá lớn, chúng ta không thể đưa hết mỗi lần 1 epoch để huấn luyện. Buộc lòng chúng ta phải chia nhỏ 1 epoch ra thành các nhỏ hơn.

số lượng mẫu dữ liệu trong một batch được gọi là batch size.

Trong quá trình huấn luyện chúng ta thường sử dụng nhiều hơn 1 epoch: để có thể tối ưu được model của chúng ta sau quá trình huấn luyện. Mục đích nhầm thu được model tối ưu nhất.

**Steps\_per\_epoch:**

**Step per epoch = số lượng dữ liệu tập huấn luyện / batch size**

Chúng ta có thể hiểu steps per epoch: là số lượng batch để hoàn thành 1 epoch.

Nếu số lượng dữ liệu tập huấn luyện chia cho batch size là lẻ thì dữ liệu của batch cuối cùng sẽ ít hơn các batch trước đó.

Số lượng dữ liệu trong một batch là: batch size.