1. Trong thực tế, tập dữ liệu **phải** được chia thành những tập nào?

**Trả lời:**

* Sử dụng **3** tập: Training, validation & testing.

1. Thêm nhiều **features.**
2. Chuẩn hóa dữ liệu có ảnh hưởng gì đến **Regularization**?
3. Một mô hình bị **High Variance.**

* Thêm training examples vào thì **chắc chắn loss** chỉ **tăng** hoặc **không đổi**.
* Khi tăng kích thước của tập training đủ lớn thì **loss** tăng rất **chậm** vì mô hình đã **tới hạn**. Dù có thêm **examples** & train tiếp thì mô hình **gần như không** thay đổi. Do dữ liệu nếu khảo sát đúng luôn có tính **xu hướng**.
* Khi **m** tăng dần đến vô cùng thì xu hướng của **2** hiện tượng Overftting & Underfittinglà **giống nhau**.

1. Yếu tố nào quyết định **Overfitting/Underfitting**?