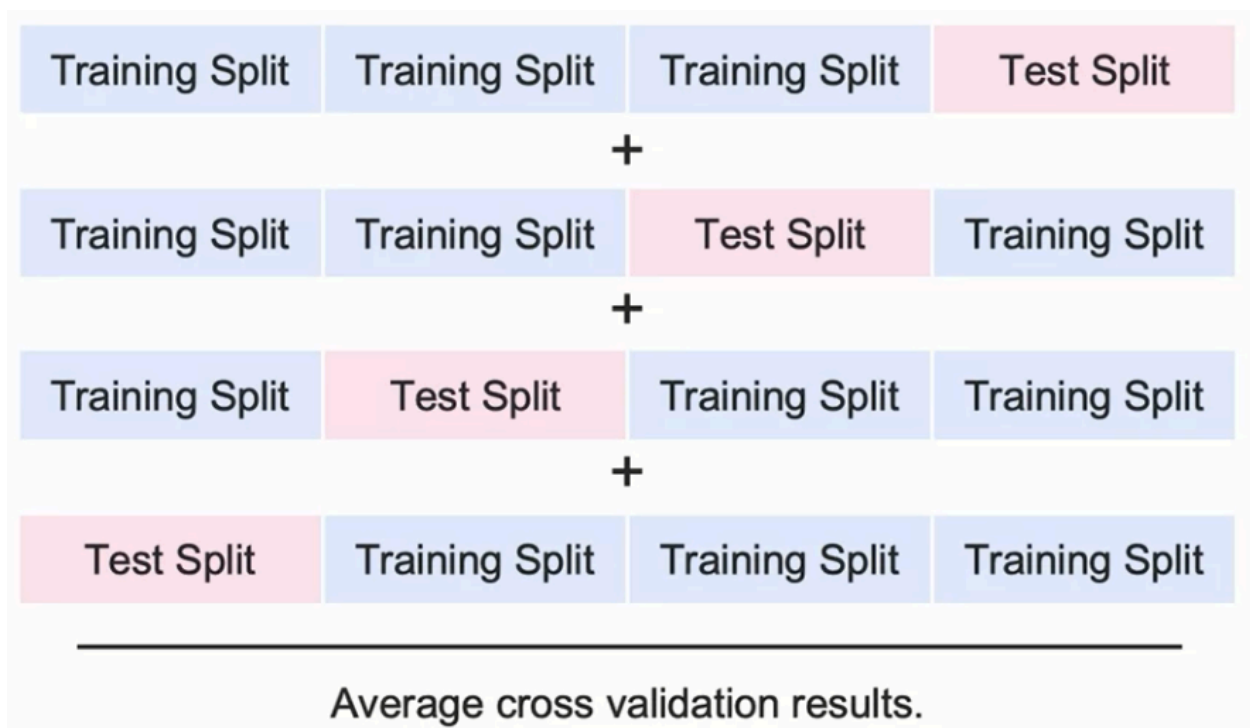


# Course2\_Module3

## Cross Validation

### Cross Validation

- Vấn đề: chia 1 lần train/test chưa chắc phản ánh đúng hiệu suất mô hình.
- Giải pháp: dùng nhiều lần chia để đánh giá trung bình → **cross validation**.



### Các phương pháp phổ biến

#### 1. k-Fold Cross Validation

- Chia dữ liệu thành k phần (folds).
- Luân phiên dùng 1 fold để test, k-1 fold còn lại để train.
- Lấy kết quả trung bình.

#### 2. Leave-One-Out Cross Validation (LOOCV)

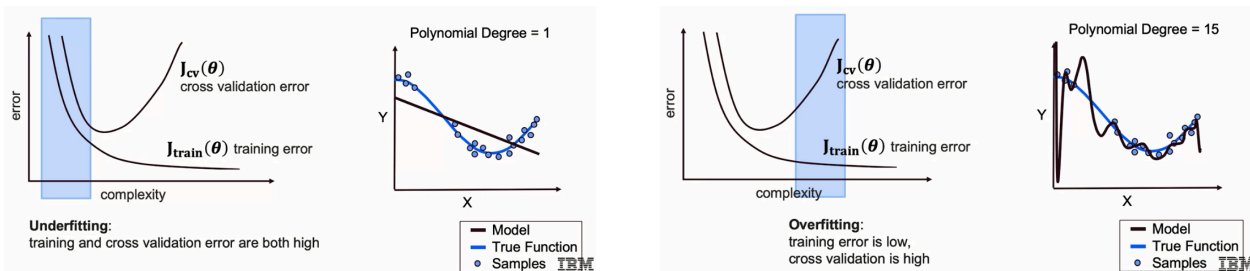
- Mỗi lần chỉ bỏ lại 1 điểm dữ liệu để test, phần còn lại để train.
- Lặp lại cho đến khi tất cả điểm đều được test.

### 3. Stratified Cross Validation

- Giữ nguyên tỉ lệ phân bố của nhãn (labels) trong từng fold.
- Đặc biệt hữu ích với dữ liệu mất cân bằng (ví dụ: spam vs. non-spam).

## Model Complexity vs. Error

- **Bias–Variance Tradeoff:**



- Mô hình đơn giản → bias cao, variance thấp.
- Mô hình phức tạp → bias thấp, variance cao.
- **Cross validation** giúp chọn mức độ phức tạp tối ưu, tránh cả underfit và overfit.