

KHAI PHÁ DỮ LIỆU

Trường Đại học Nha Trang
Khoa Công nghệ thông tin
Bộ môn Hệ thống thông tin
Giáo viên: TS.Nguyễn Khắc Cường

CHỦ ĐỀ 4

PHÂN LỚP (Đánh giá mô hình)

Đánh giá mô hình

- Datasets có sẵn tập huấn luyện, có sẵn tập kiểm tra:
 - Dùng tập huấn luyện để xây dựng model
 - Dùng tập kiểm tra để đánh giá model
- Datasets có sẵn tập huấn luyện, không có tập kiểm tra:
 - K-fold validation:
 - Chia tập huấn luyện thành k phần bằng nhau
 - Dùng k-1 phần để xây dựng model
 - Dùng phần còn lại để kiểm tra
 - Lặp lại k lần → tính trung bình
 - Hold-out validation:
 - Lấy ngẫu nhiên 2/3 phần dataset để xây dựng model
 - Dùng 1/3 phần còn lại để kiểm tra
 - Lặp lại k lần → tính trung bình

Đánh giá mô hình

- Một số metrics đánh giá model
 - Confusion matrix for two-class

Predicted Class	Actual Class	
	Class 1	Class 2
Class 1	A: True +	B: False +
Class 2	C: False –	D: True –

- Metrics đánh giá sử dụng thông tin của matrix trên

Evaluation Metrics	Computation Using Confusion Matrix
True positive rate (TP)	$TP = A/(A + C)$
False positive rate (FP)	$FP = B/(B + D)$
Sensitivity (SE)	$SE = TP$
Specificity (SP)	$SP = 1 - FP$
Accuracy (AC)	$AC = (A + D)/(A + B + C + D)$
Recall (R)	$R = A/(A + C)$
Precision (P)	$P = A/(A + B)$
F -measure (F)	$F = 2PR/(P + R)$

Đánh giá mô hình

Q / A