



CLASSIFICATION

SYRACUSE UNIVERSITY
School of Information Studies

WHAT IS DATA MINING?

“Non-trivial extraction of implicit, previously unknown and potentially useful information from data.” – Gregory Piatetsky-Shapiro, founder of kdnuggets.com

In support of decision-making



TYPICAL DATA MINING TASKS

Classification

Clustering

Association rule mining

WHAT IS CLASSIFICATION?

Given some predefined categories,
assign objects to one or more categories.

Is this fruit apple, orange,
strawberry, or pear?



?



CLASSIFICATION VS. REGRESSION

The **prediction** output is different.

Classification outputs **categorical** decisions (e.g., spam or regular e-mails).

Uses machine-learning techniques

Taught in this class

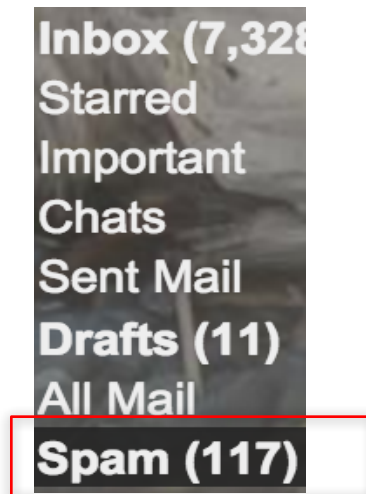
Regression outputs **numeric** values (e.g., stock price, temperature).

Uses regression techniques

Taught in statistics class

CLASSIFICATION IN EVERYDAY LIFE

Gmail identifies spam e-mails from regular ones.



Gmail categorizes regular e-mails into Primary, Social, Promotion, Updates, etc.

