ASSOCIATION RULES LEARNING

ASSOCIATION RULE LEARNING

LET'S SAY YOU WORK AT AN ECOMMERCE COMPANY AS A CATEGORY MANAGER

YOU ARE IN CHARGE OF SELLING MOBILE ACCESSORIES - THINGS LIKE CELLPHONE CASES, CHARGERS ETC

YOUR JOB IS TO SELL A LOT OF STUFF, AND AT PRICES AS HIGH AS POSSIBLE, AND SPEND AS LITTLE AS POSSIBLE ON MARKETING

WHAT IF YOU COULD FIGURE OUT, SOMEHOW, THAT FOLKS WHO BOUGHT ADAPTERS AND EARPLUGS WERE MORE LIKELY TO BUY CELLPHONE CHARGERS -

THAT INFORMATION COULD REALLY HELP - YOU COULD PERHAPS "BUNDLE" ADAPTERS AND CELLPHONE CHARGERS, OR DISPLAY PROMOTIONAL PRICING, OR OFFER QUANTITY DISCOUNTS

{Adapter, Earmuffs} -> {Cellphone Charger}

IDENTIFYING RULES OF THIS SORT IS EXACTLY WHAT ASSOCIATION RULE LEARNING IS ALL ABOUT

{Adapter, Earmuffs} -> {Cellphone Charger}

IDENTIFYING RULES OF THIS SORT IS EXACTLY WHAT **ASSOCIATION RULE LEARNING** IS ALL ABOUT

THE BASIC OBJECTIVE OF THIS TECHNIQUE IS TO UNCOVER INTERESTING RELATIONSHIPS BETWEEN VARIABLES IN LARGE DATASETS

FOR INSTANCE, SAY THAT 1% OF ALL TRANSACTIONS INVOLVE BUYING BOTH AN ADAPTER AND EARMUFFS

0.9% OF TRANSACTIONS INVOLVE BUYING AN ADAPTER, EARMUFFS, AND A CELLPHONE CHARGER

AND 3% OF ALL TRANSACTIONS INVOLVE THE PURCHASE OF A CELLPHONE CHARGER

INTUITIVELY, WE CAN SEE WE HAVE STUMBLED UPON SOMETHING, BECAUSE WHEN FOLKS BUY AN ADAPTER AND EARMUFFS, 90% OF THE TIME, THEY ALSO BUY A CHARGER..

THIS IS EXPLICITLY MEASURED USING A CONCEPT CALLED **LIFT OF A RULE**

$$Lift(X \rightarrow Y) = \frac{\text{proportion of transactions with both X and Y}}{\text{proportion of transactions with X} \times \text{proportion of transactions with Y}}$$

X = {ADAPTER, EARMUFFS}

Y = CELLPHONE CHARGER

$$Lift(\{Adapter, Earmuffs\} \rightarrow Cellphone\ charger) = \frac{0.9\%}{0.1\% \times 3\%} = 3$$

FOLKS ARE 3 TIMES MORE LIKELY TO BUY A CHARGER ALONG WITH A COMBINATION OF (AN ADAPTER, EARMUFFS) THAN IF THESE 3 ITEMS WERE INDEPENDENT

A MACHINE-LEARNING BASED IMPLEMENTATION OF ASSOCIATION RULE LEARNING

WOULD INVOLVE FINDING A LARGE NUMBER OF SUCH RULES EFFICIENTLY, AND ALSO SUBJECTING THESE RULES TO TESTS OF CONVICTION (HOW LIKELY IS THE RULE TO BE WRONG?)

THE ORIGINAL ACADEMIC PAPER THAT PROPOSED SUCH RULES IS A CLASSIC - IT HAS APPARENTLY BEEN CITED ABOUT 18,000 TIMES