SELEZIONE DI ATTRIBUTI

SELEZIONARE, TRA TUTTI GLI ATTRIBUTI A DISPOSIZIONE, UN SOTTOINSIEME SIGNIFICATIVO PER L'ATTIVITA' DI ESTRAZIONE DI CONOSCENZA

- B I DATABASE CONTENGONO MOLTI ATTRIBUTI INSERITI PER MOTIVI DIVERSI DAL DM
- ß ALCUNI ATTRIBUTI POSSONO ESSERE IRRIVELANTI ED INGANNARE L'ALGORITMO A CAUSA DI INTERAZIONI NON DESIDERATE
- B ALCUNI ATTRIBUTI POSSONO ESSERE MOLTO RUMOROSI E "CONFONDERE" L'ALGORITMO

(COURSE OF DIMENSIONALITY)

LA SELEZIONE DI ATTRIBUTI

- ß AUMENTA L'ACCURATEZZA DEL MODELLO
- B AUMENTA L'EFFICIENZA DEL PROCESSO COMPLESSIVO DI DM

SELEZIONE MANUALE DI ATTRIBUTI

L'UTENTE SPECIFICA QUALI ATTRIBUTI ELIMINARE

VANTAGGI:

- ß SI SFRUTTA LA CONOSCENZA DI DOMINIO DI UN ESPERTO
- **B** SI RIDUCE IL TEMPO DI COMPUTAZIONE

SVANTAGGI:

- **B** MANCATA OTTIMIZZAZIONE DEI DATI
- B DIFFICOLTA' NELL'ELICITAZIONE DELLA CONOSCENZA DI UN ESPERTO
 - o INTERAZIONE TRA ATTRIBUTI
 - o ATTRIBUTI APPARENTEMENTE IRRILEVANTI POSSONO PORTARE ALLA SCOPERTA DI CONOSCENZA NUOVA
- **B** METODO SOGGETTIVO

SELEZIONE AUTOMATICA DI ATTRIBUTI

SPAZIO DI RICERCA MOLTO GRANDE: $2^{|N^{\circ} \text{ ATTRIBUTI}|}$

CLASSIFICAZIONE APPROCCI PRINCIPALI:

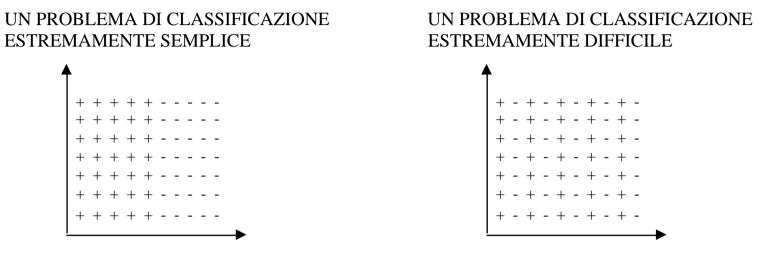
- **B** INDIPENDENTE DAL TIPO DI ALGORITMO DI DATA MINING (FILTER APPROACH)
 - o ANALISI DELLE COMPONENTI PRINCIPALI
 - o ANALISI CORRELAZIONI
 - INDUZIONE DI REGOLE
 - o RELIEF ALGORITHM
 - o METODI BASATI SULLA TEORIA DELL'INFORMAZIONE
- **B** OTTIMIZZATI PER IL DATO ALGORITMO (WRAPPER APPROACH)
 - 1) EURISTICA PER LA SELEZIONE DEL SOTTOINSIEME DI ATTRIBUTI
 - 2) FUNZIONE DI VALUTAZIONE (ACCURATEZZA)

FILTER VS WRAPPER APPROACH

- β FILTER APPROACH PIU' EFFICIENTE
- **B** WRAPPER APPROACH PIU' EFFICACE

ESTRAZIONE DI ATTRIBUTI

"NESSUNA TECNICA FUNZIONA BENE SE LE CARATTERISTICHE MISURATE NON SONO ADEGUATE" (PROVATE A TROVARE UN MODO DI PREDIRE IL TEMPO A PARTIRE DALL'ALTEZZA DEL VOSTRO VICINO..!)



IN GENERALE

- ß PROBLEMI DI CLASSIFICAZIONE SEMPLICE ASSOCIATI A RAPPRESENTAZIONI DEI DATI ASTRATTE, "BASATE SU CONOSCENZA"
- ß PROBLEMI DI CLASSIFICAZIONE DIFFICILE ASSOCIATI A RAPPRESENTAZIONI DEI DATI DI BASSO LIVELLO

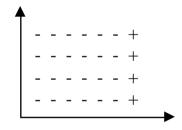
ESEMPIO

UN SUPERMERCATO VUOLE PREDIRE SE LA QUANTITA' DI VENDITE DI UN CERTO PRODOTTO SARA' ALTA (+) O BASSA (-), SULLA BASE DELLA CONOSCENZA DELLA SOLA DATA

RAPPRESENTAZIONE DELLA DATA COME GIORNO DEL MESE

_ - - - - - + - - - - - + - - - - - + + - - - - - + +

RAPPRESENTAZIONE DELLA DATA COME GIORNO DELLA SETTIMANA



IN MOLTI CASI IL SUCCESSO DELL'INDUZIONE DI REGOLE DIPENDE PIU' DALL'USO DI RAPPRESENTAZIONI "KNOWLEDGE BASED" CHE DALL'USO DI UN ALGORITMO INTELLIGENTE

L'USO DI CONOSCENZA DI DOMINIO RIDUCE L'AUTONOMIA DEL SISTEMA, ANDANDO CONTRO LO SPITIRITO DEL PROCESSO DI KD E RIDUCENDO LA POSSIBILITA' DI SCOPRIRE CONOSCENZA VERAMENTE NUOVA E INTERESSANTE