

**POLITECNICO DI TORINO**  
**ESAMI DI STATO PER L'ABILITAZIONE**  
**ALLA PROFESSIONE DI INGEGNERE**  
**I SESSIONE - ANNO 2004**  
Ramo: INFORMATICA                      Tema n. 2

Premessa.

Molti fenomeni fisici vengono rappresentati mediante un insieme di misure di alcune caratteristiche distintive (*features*), organizzate in un vettore. Si pensi, ad esempio, ai fenomeni atmosferici, descritti dall'insieme di temperature, pressione, umidità, ecc., o ad un segnale elettrico stazionario, descritto dai punti dello spettro, e così via. Questi vettori possono essere considerati dei punti in uno spazio euclideo multidimensionale.

In molte applicazioni, piuttosto che trattare i vettori veri, si preferisce sostituirli con dei "rappresentanti" opportunamente scelti. Il procedimento, detto quantizzazione vettoriale, è descritto qui di seguito.

- I. All'inizio tutti i vettori sono attribuiti ad un unico cluster. Si calcola l'unico rappresentante come valor medio di tutti i vettori;
  
- II Per ciascun rappresentante finora trovato:
  - A. si crea una perturbazione (casuale o meglio ancora lungo la direzione di massima dispersione) in modo da determinare due nuovi rappresentanti;
  - B. si assegna ciascun vettore del rappresentante padre ad uno dei due figli, secondo il criterio della minima distanza;
  - C. si ricalcolano i nuovi rappresentanti come il vettore medio dei vettori assegnati a ciascun figlio: saranno quelli i nuovi centri di cluster (si abbandonano i vettori "perturbati").
  
- III. Se *non* è soddisfatto il "criterio di fine", si va a II.

Alla fine i vettori rappresentanti, detti *codeword*, sono etichettati con un intero da 0 a  $2^N$ , dove N è il numero di bit utilizzato per codificare i vettori.

Il "criterio di fine" può essere legato:

- 1) al raggiungimento del numero di bit N stabilito a priori per la codifica;
- 2) al superamento di una soglia minima di "distorsione" accettata, se si è definito un opportuno indice di distorsione riferita all'applicazione che si sta trattando. Nel caso più semplice, si può imporre che la distanza media dei vettori rappresentati dal vettore rappresentante sia sotto una data soglia.

NB: sostituire i vettori veri con i rispettivi rappresentanti può essere considerato come una *distorsione* dello spazio vero; ovvero, se si stanno trattando sequenze di vettori, ciascuno dei quali è un punto nello spazio, al *pattern* (traiettoria) vero si sostituisce un pattern distorto, quello che è costretto a passare per i punti (vettori) rappresentanti piuttosto che per i punti veri. In quest'ultimo caso il numero di bit N utilizzato per la codifica non è stabilito a priori, ma è quello trovato.

(segue)

Campi di utilizzo:

- 1 ) dove serve una riduzione di dimensionalità (compressione di banda per qualsiasi applicazione)
- 2) dove serve passare da variabili continue a variabili simboliche (associando etichette ai codici)

Tema.

In un file di tipo testo sono memorizzati un insieme di vettori da utilizzare per la quantizzazione vettoriale. Nella prima riga un numero intero indica la dimensionalità dei vettori. Nelle righe successive ci sono gli elementi dei vettori espressi come numeri reali, separati da almeno uno spazio (blank) o da un RETURN (<CR>). Il numero di vettori può superare le migliaia.

Realizzare un programma in linguaggio C che attui l'algoritmo di quantizzazione vettoriale sopra descritto.

Il nome del file dei dati deve essere introdotto da tastiera o passato come parametro.

Costituiscono altri parametri del programma:

- la dimensionalità dei vettori effettiva da trattare (minore o uguale a quella descritta nel file);
- la modalità di terminazione, se basata sul numero di bit delle codeword oppure sulla distanza media dei vettori dalla rispettiva codeword. Nel primo caso occorre fornire il numero di bit, nel secondo la soglia.

Descrivere nel dettaglio ma in modo sintetico le strutture dati e gli algoritmi utilizzati nella manipolazione dei dati per realizzare il progetto. Quanto non stabilito espressamente deve essere scelto dall'esaminando e sarà oggetto di valutazione.