

Sequential Patterns

Occorrenze : $\langle\{A,B\}\rangle \langle\{A,C\}\rangle \dots$

- min_gap = 1 --> distanze > di uno: $\langle 2,3 \rangle$ NO
- max_gap = 2 --> distanze \leq a 2: $\langle 2,5 \rangle$ NO
- max_span = 3 --> distanza \leq a 3: $\langle 2,3,5 \rangle$ NO

GSP: ab --> a --> b

- 1 - Genero gli 1-itemset e valuto il supporto;
- 2 - Genero tutte le sequenze dagli itemset validi e valuto il supporto;
- 3 - Genero le 3-sequenze dal passo due, valuto se per ogni 3sequenza, togliendo uno alla volta un membro, ha supporto $>$ sup_min, se si la tengo altrimenti la elimino. Successivamente valuto il supporto delle 3-sequenze rimaste;
- 4 - Continuo come al passo 3: genero, pruning, valuto supporto.

Serie temporali

KNN: distanza euclidea tra una serie e tutte quelle del training e prendo i k più bassi.

DTW: matrici enormi. Scelgo quella con vertice altro-destra minore e seleziono il cammino con minor costo. La matrice scelta determina la classe.

Classificazione

KNN: scelgo i k-vicini con minor distanza

Dati continui: distanza euclidea;

Dati categorici:

- 1) matrici attributi/classi;
- 2) $d(\text{alto}/\text{basso}) = |(\text{alti classe SI}/\text{tot alti}) - (\text{bassi classe SI}/\text{tot bassi})| + |\text{stessa cosa per classe NO}|$;
- 3) somma distanze al quadrato.

Decision Tree:

- a) Controllare se le regole possono essere semplificate (magari un solo attributo a sinistra) attraverso il $PE = (Training_ERROR + 0.5)/N$. La regola con PE minore è quella migliore;
- b) Costruire le matrici di confusione per i modelli e valutare statistiche sia sul training che sul test set:
 - o/c | YES NO | Accuracy: $(TP+TN)/\text{tutti}$
 - YES | TP FN | Precision: $TP/(TP+FP)$
 - NO | FP TN | Recall: $TP/(TP+FN)$

Bayesian Model: scelgo la classe dell'elemento con P maggiore

Dati continui: media, varianza, funzione (per entrambe le classi);

Dati categorici: Produttoria di $P(X|YES) * P(YES)$

dove $P(X|YES) = (n.\text{con } X) / (n.\text{tot } YES)$

alla fine: matrice di confusione e statistiche.

Dati continui E categorici come i casi precedenti e moltiplicazione tra le varie probabilità.



By **perascotta**

cheatography.com/perascotta/

Not published yet.

Last updated 12th April, 2015.

Page 1 of 1.

Sponsored by **Readability-Score.com**

Measure your website readability!

<https://readability-score.com>