**ARGOMENTI DEGLI ESAMI**

03 /11 /20

tf-idf con documenti lunghi e cosine similarity (ariccia) 2

Naive bayes

Malattie: markow chain, stationary distribution

Esame 18/06/18

Tf-idf (es: data mining), df, posting lists 2

Svm (grafico decision boundary)

Grafo pagerank con presonalization vector

23/01/19

Tf-idf. Classificazione documenti cin cosine similarity (tipo 10 volte)

Almeno una ragione per cosine similarity megli di euclidean distance (tipo 10 volte)

Hits, algoritmo, esempio con nodi hub score=0 (1 volta)

Pagerank, small citation network, differenza tra pagerank e papers’ scientific impact (1 volta)

Rank respect to authoritativeness and to closeness (1 volta)

21/09/2017

Stemming, Skippointers AND NOT, Biword index (5 volte)

Zipf’s law, Byte encoding (2 volte)

Teleporting prob. Importance, Equazione del grafo, risouzione con a/2 (praticamnete sempre)

Explain k-means algorithm (algoritmo), unlucky assignment k means solution, why k-means converge (tipo 7 volte)

18/03/2021

Postings lists: still interesction O(x+y) for Brutus AND NOT Ceasar. And Algorithm (molte volte)

Ricerca parole del tipo m\*nchien e tecnica usata (2 volte)

Tf-idf(data mining) (2 volte)

Document cosine similarity (ariccia) (2 volte)

11/02/19

a AND b algorithm (spesso)

svm grafico (3 volte)

grafo con equazioni generali pagerank, pagerank con personalized vector (praticamente sempre)

12/04/19

a OR b algorithm (qualche volta)

Svm vs knn con grafo a spirale (1 volta)

Grafo personalization vector, compute personalized pagerank senza risolvere il sistema con pers, vector = {0.5,0.5,0,0} (3 volte)

11/01/12

Pagerank, importance of teleporting probability, con grafo 7 nodi a= ½ (2 volte)

Compression yte encoding and gamma encoding (3 volte)

a AND b AND NOT c algorithm (5 volte)

Esempio pratico comparazione postings lists normale e con skip (2 volte)

26/01/12

Stemming, utilità skip pointers per a OR b, biword index (qualche volta=

Dimostra che euclidea dist con vettori normalizzati ha stesso ordine di prossimita di cosine sim. (praticamente sempre ma non troviamo la seconda ragione)

Grafo. Equazioni, a=1/2, prova che per ogni nodo il pagerank è almeno a/N (spessissimo)

Knn, spiega knn classification con inverted list ( tipo 10 volte)

K-means alg, esempio con i pallini (tipo 10 volte)

02/02/2012

Dove sono piu utili gli skip PS: Termini frequenti = postings lists lunghe (2 volte)

a AND b alg (spesso)

Grafo equations, a= ½ . Markow chain con beta in Vo e markov chain con back button ??????????????????? (1 volta)

Tf-idf formula e spiegazione. Comparazione 3 documenti (5 volte)

2 ragioni euclidean peggio di cosine (spessissimo)

NNNNRRR precision, recall, curvam interpolated precision (spessissimo non sappiamo fare interpolazione)

20/02/2013

Compression byte encoding, gamma (3 volte)

Grafo 7 nodi, importacne teleporting, eqazioni e a=1/2 (2 volte)

Tf-idf waiting equation. Classificazione 3 documenti con cosine simil (5 volte)

2 ragioni cosine meglio euclidea (spessissimo)

Knn alg, knn classification con inveted list, spiega k-means, disegna esempio con pallini k-means (7 volte)

06/07/13

x AND NOT y naive evaluation, why is expensive, posting merge list alg (3 volte)

struttura per rispondere “in un piatto poco cupo, poco pepe cape”. Dai un esempio di indice (2 volte)

NNNRR precision, recall curva (spesso)

Tf-idf equation, classificazione 3 documenti con cosine (5 volte)

2 ragioni cosine meglio euclidea (spessissimo)

Importance of teleporting prob. Grafo equazioni e prob a= ½ (praticamente sempre)

Descrivi assunzioni naive bayes, computa i coefficienti di boolean classifier di 4 documents (2 volte)

Classifica la query toyota jeep (2 volte)

06/09/2012

Stemming, pointers not useful per x OR y, biword index (spesso)

Normalize vector euclidean uguale a cosine (what if not normalized) (3 volte)

Grafo, importance telep, a= ½ , dimostr pagerank almeno a/N (praticamente sempre)

Knn alg, knn class, k-means, esempio k means con pallini (spesso)

07/20/12

X AND y AND NOT z naive, piatto poco cupo (2 volte)

3 tier-index, Cluster prunning (1 volta)

RNNNR prec, rec, F1 score, curva, interpolation ( tipo 7 volte)

Grafo, a=0, a=1/2 (praticamente sempre)

09/09/12

Skip pointers useful, merging a AND b con skip (2 volte)

Grafo, importance telep, a= ½, pagerand almeno a/N (3 volte)

Tf-idf weigthing classificiazione 3 doc con cosine ( 5 volte)

Almeno 2 ragioni cosine meglio euclidean (7 volte)

24/01/2013

A AND NOT b, posting merge. Queries x/3 y/4 z con k proximity operator. Dai un esempio di tale index (2 volte ma non sappiamo fare)

RRRNNN prec, recall, curve (7 volte)

Descrivi external memory alg for power iteration method, why fast convergence, esegui l’algoritmo sul grafo con inizial state (1,0,0) e a=1/2 (2 volte)

Grafo, equazioni, a= ½ (praticamente sempre)

05/07/2013

Pseudocode intersection a and b and c. (spessissimo)

Cosidera la query, quante comparazioni per fare intersezione ( 2 volte)

Modifica e trova pseudo per a and b and not c ( 3 volte)

Byte e gamma encoding and decompress (2 volte) e decompress 1 volta

Norm vectors euclidean =cosine. Computare k nearest documents with cosine (3 volte)

Graph a=0, a=1/2 (spessissimo)

Naive bayes classifier, compute boolean classifier di 4 ocumenti, classify toyota jeep (2 volte)

11/09/13

Uguale al precedente ma l’ultimo esercizio è il k-means con esempio dei pallini ( svariate volte)

27/01/14

A and y and not z alg. Numero di comparazioni per le posting lists ( 3 volte)

X and not y. Len x= 1 len y=n upper bound di comparazioni (1 volta)

RNNNR prec, rec, curva, interpolazione (7 volte)

External memory alg power iteration method, (2 volte)

execute algorithm vettore iniziale (1,0,0) a=1/3 3 iterazioni. Desciriv il set di equazioni per pagerank (1 volta)

naive bayes assumpion in bag of words model, compute the coefficient of boolean classifier without smoohing. Classify lion mountain safari (2 volte)

19/01/15

Stemming, Skip useful x AND NOT y, biword index ( 3 volte)

NNNRRR prec, rec, F1, curva (7 volte)

Grafo: personalized (1,0,0,0), (0,1,0,0), explain how to calculate (0.5,0.5,0,0) senza risolvere sistema (3 volte)

Tf-idf formula, classificazione 3 docs con cosine, 2 ragioni cosine meglio euclidean ( spesso)

16/04/15

Document ranking in vectorial model, tf-idf forumla. Top-K algoritmo (1 volta)

NNNRR prec, rec, ci deve essere un break-even point tra precision e recall? (1 volta)

Grafo equazioni, at every step con prob a il random surfer jumpa a v1. Il processo risultante è sempre markov chain? (1 volta)

Naive bayes bag of words. Compute the boolean class without smoothing 4 documenti, query lion mountain safari (2 volte)

20/07/15

Stemmin, skip a and not y, biword index (3 volte)

RRNNN prec rec F1, curva (spesso)

Graph, pers vect (1,0,0,0), ((0,1,0,0), computa (0.5,0.5,0,0) senza risolvere sistema (3 volte)

Tf-idf forula, classificazione 3 doc con cosine, 2 ragioni cosine meglio euclidean (spesso)

20/10/15

A and not b si puo runnare intersezione O(x+y)? E per a OR NOT b (1 volta)

Posting merge a and not b (spesso)

Zipf’s distribution( roba brutta della prob) (2 volte)

Compression byte encoding (3 volte)

Grafo importance telep, equazioni a=1/2 (praticamente sempre)

K-means, pallini (7 volte)