

B003725 Intelligenza Artificiale (2017/18)

Studente: Manuel Salamino (5992236) — <2018-03-13 Tue>

Elaborato assegnato per l'esame finale

Istruzioni generali

Il lavoro svolto sarà oggetto di discussione durante l'esame orale e dovrà essere sottomesso per email due giorni prima dell'esame, includendo:

1. Links ad un repository **pubblico** su [github](#), [gitlab](#), o [bitbucket](#), dove reperire sorgenti e materiale sviluppato. In alternativa è accettabile allegare un file zip.
2. Un file README che spieghi:
 - come usare il codice per riprodurre i risultati sottomessi
 - se vi sono parti del lavoro riprese da altre fonti (che dovranno essere **opportunamente citate**)
3. Una breve relazione (massimo 4 pagine in formato pdf) che descriva il lavoro ed i risultati sperimentali. Non è necessario ripetere in dettaglio i contenuti del libro di testo o di eventuali articoli, è invece necessario che vengano fornite informazioni sufficienti a *riprodurre* i risultati riportati.

Importante: negli eventuali allegati va evitato di inserire files eseguibili (inclusi files .jar o .class generati da Java), al fine di evitare il filtraggio automatico da parte del software antispam di ateneo!

Naive-Bayes e curve ROC

Nella prima parte di questo esercizio si implementa, in un linguaggio di programmazione a scelta, l'algoritmo Naive Bayes come esposto in classe e descritto in [McCallum & Nigam 1998](#) (nella versione Bernoulli). Si sviluppa quindi del codice per calcolare curve ROC come descritto in [\(Davis & Goadric, 2006\)](#) usando come score la distanza dall'iperpiano di separazione.

Nella seconda parte, si applica il codice al data set [20 newsgroups](#), scegliendo una classe come positiva e le altre come negative (ripetendo per almeno tre classi).