

به نام خدا

گزارش فاز دوم پروژه پایانی درس یادگیری ماشین

استاد درس: دکتر محمدحسین رهبان

پویا خانی

تابستان ۱۴۰۰

بخش ۱-ب : در این بخش به عنوان روش سوم از روش birch استفاده کردم. معیار هایی که استفاده کردم هم همان تابع analysis فاز اول، و معیار rand_index و purity است. نتایج به شرح زیر است:

Model and parameters	Accuracy	Rand_index	Purity
K-means with 2 cluster & kmeans++ init and 10 n_init	0.6237	0.5306	0.6237
GMM with 2 cluster	0.6274	0.5324	0.6274
Birch with 2 cluster & threshold 0.079	0.6261	0.5318	0.6261

بخش ۱-پ : در این قسمت از الگوریتم birch با ۳ کلاستر استفاده کردم. با خوشه بندی داده ها و نمایش جملات مربوطه، دیدم که داده های درون خوشه با برچسب صفر ، جملات با مضمون مثبت و تمجید است و داده های با برچسب ۱ ، جملات خنثی و توصیفی است که نظر مثبت یا منفی ندارند و فقط درباره موضوعی مطلبی نوشته اند و جملات با برچسب ۲ ، جملات منفی و با مضمون انتقاد است.

بخش ۲) در این بخش، برای قسمت الف ، از بردارسازی BOW و MLP با ۶ لایه مخفی به ترتیب با ۵۰۰ و ۲۵۰ و ۱۰۰ و ۵۰ و ۱۰ و ۵ نورون با تابع فعالسازی tanh استفاده کردم که دقت ۷۲ درصد بهم داد. حال با استفاده از بهترین مدل فاز ۱ یعنی مدل logistic regression با بردار سازی word2vec ، با لود کردن مدل word2vec فاز اول و لود کردن مدل logistic regression مربوطه با پارامتر warm_start=True ، به دقت ۰.۷۵ درصد رسیدم. یعنی استفاده از مدل pretrain شده بهبود چند درصدی را حاصل کرد.

مابقی گزارش ها و اطلاعات در نوتبوک موجود است.