

به نام خدا

گزارش فاز دوم تمرین شماره ۲

قسمت اول:

در این قسمت Classifier نایبویز (با و بدون smoothing) پیاده‌سازی شده‌است. ۱۰ درصد از کل واحد های داده‌ای (۳ بیت در غزلیات و ۶ بند در اشعار نیمایی) به تست اختصاص داده شده‌است. در انتهای برنامه precision و recall محاسبه می‌گردد که بعد از ارتقای برنامه به نایبویز با smoothing شاهد بهبود این دو معیار خواهیم بود.

نکته: در این فاز زمان جمع‌آوری داده‌ها حذف حروف اضافه و کلمات بی‌هوده را انجام ندادیم اما نرمال‌سازی و همچنین یکسان‌سازی فعل‌ها انجام می‌شود. (در جهت تلاش برای محاسبه‌ی دقیق‌تر احتمالات)

قسمت دوم:

در این قسمت هم به صورت مشابه ۱۰ درصد داده‌ها به داده‌های آزمایشی اختصاص دادیم. سپس داده های آموزش و آزمایش را طبق ساختار موردنیاز برای vowpal wabbit درآوردیم. یک اسکریپت تحت زبان برنامه نویسی پایتون نوشته شده‌است که شامل دستورات ترمینالی لینوکس می‌باشد. در هر مرحله به ازای n های ۱ تا ۵ تست داده‌ها انجام می‌گردد و precision و recall محاسبه می‌شود. در این مجموعه داده‌ها شاهد آن هستیم که تمامی تست ها در $n = \{1,2,3,4,5\}$ درست پیش‌بینی می‌شوند!