



بازیابی پیشرفته اطلاعات

نیم سال اول ۱۴۰۲-۰۳

مدرس: دکتر حمید بیگی

۴ آذر

زمان: ۲۵ دقیقه

کوییز سوم (۱۰۰ نمره)

سوال ۱: تخمین احتمال تولید کوثری (۴۰)

در نظر بگیرید که سه داکيومنت زیر ($d1, d2, d3$) را در Collection داریم. به کمک روش هموارسازی Jelinek-Mercer smoothing احتمال تولید کوثری را برای هر داکيومنت برای کوثری داده شده محاسبه کنید و در انتها سندها را بر اساس این احتمالات رتبه بندی کنید. λ را برابر $1/4$ در نظر بگیرید.

d1: Shakespeare was a renowned playwright and wrote poets
 d2: the famous playwright William wrote the tragedy of Hamlet
 d3: Works of Shakespeare are celebrated worldwide
 query: Shakespeare wrote Hamlet

سوال ۲: naive bayes (۵۰)

با توجه به جدول زیر به سوالات الف-د پاسخ دهید.
 الف) یک طبقه بندی چند جمله ای با استفاده از روش naive bayes را روی داده های جدول تخمین بزنید.
 ب) طبقه بندی به دست آمده را روی داده تست اعمال کنید.
 ج) یک طبقه بندی برنولی با استفاده از روش naive bayes را روی داده های جدول تخمین بزنید.
 د) طبقه بندی به دست آمده را روی داده تست اعمال کنید.
 نکته: در جدول زیر داک ۱، داک ۲، داک ۳ و داک ۴ مربوط به داده آموزش است و داک ۵ مربوط به داده تست است.

آیدی داکيومنت	کلمات موجود در داکيومنت	آیا مهم است؟
داک ۱	باران تند	بله
داک ۲	برف تند شد	بله
داک ۳	حيوان آمد	خير
داک ۴	آمد پلنگ تند	خير
داک ۵	تند تند آمد	؟

سوال ۳: kNN vs SVM (۱۰)

روش k-Nearest Neighbors (kNN) با روش Support Vector Machines (SVMs) از نظر مرزهای تصمیم گیری (decision boundaries) مقایسه کنید و تفاوتها را ذکر نمایید.