



دکتر طاهائی
دانشکده ریاضی و علوم کامپیوتر

مقدمه ای بر داده کاوی
نیمسال اول ۱۴۰۳-۱۴۰۴
گردآورندگان: نرگس قنبری- امیرعلی معروفی

فاکتورگیری نامنفی ماتریس

سوال ۱

با توجه به n نقطه داده در فضای d بعدی، می‌توانیم آن‌ها را به عنوان یک ماتریس $n \times d$ به نام X نمایش دهیم، که در آن سطرهای X برای نقاط داده مختلف و ستون‌ها برای ویژگی‌های مختلف هستند. الگوریتم NMF که در درس با آن آشنا شدید، تقریب ماتریس با رتبه پایین را با این قید انجام می‌دهد که هر دو ماتریس با رتبه پایین باید غیرمنفی باشند. در این تمرین، خواهیم آموخت که خوشه‌بندی k -Means نیز می‌تواند به عنوان یک فرم خاص از تقریب ماتریس با رتبه پایین دیده شود. بر اساس این موضوع، ما شباهت‌ها و تفاوت‌های بین این دو روش خوشه‌بندی را مقایسه خواهیم کرد.

اثبات کنید خوشه‌بندی K -Means می‌تواند به عنوان یک فرم خاص از تقریب ماتریس با رتبه پایین (low-rank approximation) دیده شود. یعنی تابع هدف بهینه‌سازی K -Means معادل است با:

$$\{W^*, H^*\} = \arg \min_{W, H} \|X - HW\|_{fro}^2,$$

که در آن $\|\cdot\|_{fro}$ نرم فروبینیوس ماتریس است و H و W دو ماتریس با رتبه پایین و با قیود مناسب هستند. به طور خاص، قید اندازه H و W به چه صورت است؟ چه قیود اضافی باید بر روی H و/یا W اعمال کنیم تا تساوی بالا معادل با تابع هدف الگوریتم k -Means باشد؟