STAT 185 - Problem Set #4

NAME

This assignment covers Nonnegative Matrix Factorization and Archetypal Analysis through a mixture of theory and coding using the packages NMF and archetypes. You will need the file *pset4.Rdata* available on Canvas. The commands for loading the data assume this file, *pset4.Rmd*, and the Rdata file are in the same folder.

1. In this problem, we are going to derive the multiplicative update rules from Lee and Seung (2001) assuming NMF under the Frobinius norm

$$\|X - WH\|_F = \sqrt{\sum_{i,j} [X_{ij} - (WH)_{ij}]^2}$$

where $X \in \mathbb{R}^{N\times p}$, $W \in \mathbb{R}^{N\times k}$, and $H \in \mathbb{R}^{k\times p}$. Given small step-sizes Δ^W_{ij} and Δ^H_{jl}, a variant of the gradient descent algorithm proceeds by

   i. updating W_{ij} holding all other entries of W and H fixed, i.e.

$$W_{ij} = W_{ij} - \Delta^W_{ij} \frac{\partial}{\partial_{W_{ij}}} \left(\frac{1}{2} \|X - WH\|_F^2 \right)$$

   and iterating over $1 \le i \le N$ and $1 \le j \le k$.

   ii. updating H_{jl} holding all other entries of W and H fixed, i.e.

$$H_{jl} = H_{jl} - \Delta^H_{jl} \frac{\partial}{\partial_{H_{jl}}} \left(\frac{1}{2} \|X - WH\|_F^2 \right)$$

   and iterating over $1 \le j \le k$ and $1 \le l \le p$.

   iii. Repeating (i) and (ii) until $\|X - WH\|_F$ reaches a local minima within a desired tolerance.

   Complete parts (a) - (c) to derive the multiplicative update rule from Lee and Seung (2001).

   a. Find the partial derivative of $\frac{1}{2}\|X - WH\|_F^2$ with respect to W_{ij}.

   b. Find the partial derivative of $\frac{1}{2}\|X - WH\|_F^2$ with respect to H_{jl}.

   c. Setting $\Delta^W_{ij} = \frac{W_{ij}}{(WHH^T)_{ij}}$ and $\Delta^H_{jl} = \frac{H_{jl}}{(W^TWH)_{jl}}$ derive multiplicative updates for W_{ij} and H_{jl} by combining your answers to (a) and (b) with steps (i) and (ii) above.

2. Suppose we are interested in NMF using the *Divergence* as a measure of error. From class, the multiplicative updates are

$$W_{ij} \leftarrow W_{ij} \frac{\sum_{l=1}^p [H_{jl} X_{il} / (WH)_{il}]}{\sum_{l=1}^p H_{jl}}$$

$$H_{jl} \leftarrow H_{jl} \frac{\sum_{i=1}^N [W_{ij} X_{il} / (WH)_{il}]}{\sum_{i=1}^N W_{ij}}$$

   a. In the rank one case, show that the multiplicative updates reduce to

$$W_i \leftarrow W_i \frac{\sum_{l=1}^p X_{il}}{\sum_{l=1}^p W_i H_l}$$

$$H_l \leftarrow H_l \frac{\sum_{i=1}^N X_{il}}{\sum_{i=1}^N W_i H_l}$$

where $W_i = W_{i1}$ and $H_l = H_{1l}$.

b. Show that the iterates of (a) converge to

$$W_i = c \sum_{l=1}^{p} X_{il}, \qquad H_l = c' \sum_{i=1}^{N} X_{il}$$

for some constants $c, c' > 0$.

3. To assess the health impact of various diets on longevity, a researcher chose $N = 10$ diets. The researcher then fed diet i to N_i mice. After 6 months on the diet, the researcher counted the number of mice that were (1) healthy, (2) deceased, (3) living with cancer, and (4) living with heart disease. For simplicity assume one mouse can only experience one of the four outcomes. Let $X \in \mathbb{R}^{10\times 4}$ data matrix with counts of outcomes so that X_{ij} represents the number of mice fed diet i that experienced outcome j.

   a. Suppose we believe the outcome is *independent* of diet. Translate this into a rank assumption on X. What physical interpretation do the elements of W and H have in the nonnegative matrix factorization $X = WH$?

   b. Compute the NMF of *problem2* containing records of health outcomes using both the *Divergence* and *Frobenius* measures of error. For each method, perform 25 runs and select the outcome with the smallest residual.

   c. Assuming independence, use your best result from part (b) to estimate the probability of outcome j for $j = 1, \ldots, 4$. Output this answer as a probability vector with entries rounded to two decimals. NOTE: You may have to modify/transform the output of (b) to align with the physical interpretation in part (a).

4. Consider the 250×2 data matrix, *problem4*, comprised of 250 samples in 2-dimensional space.

   a. Generate a scatterplot of these data. Additionally, add the convex hull of these data to the plot. Indices of points on the convex hull can be found using the command *chull*.

   b. Using the command *archetype* find **the** archetype of the data in *problem4*. Generate a new plot with a scatter of the data and convex hull in black and the archetype is red.

   c. Repeat part b for $k = 1, \ldots, 8$ archetype(s) and show the successive plots in a 2×4 grid.

5. To compare *NMF* and *archetypal* analysis, consider the handwritten *digits* stored in the variable *digits* with *digits$pixels* containing one image per row and *digits$labels* containing the label of the digit.

   a. Extract the first 200 handwritten 8s from these data and save to a variable *digits8*. Find all pixels (columns) which are identically 0 over all images. Save the indices to a variable *zerocol* then remove these columns from *digits8*. (This is a necessary pre-processing step since zero columns would give rise to 0/0 terms in the multiplicative updates.)

   b. Compute 5 Nonnegative Matrix Factorizations at rank $k = 1, \ldots, 25$ using randomized initial conditions. For each k, select the fit with lowest RSS, then plot the RSS as a function of the rank k.

   c. Generate a 5×5 grid of images of the 25 features found from the final iteration of your computation from part (b). You will need to expand these feature vectors to 784-dimensions maintaining the zero rows found in part (a).

6. For comparision, consider the same analysis as problem 5 using Archetypal Analysis.

   a. Repeat 5b, using Archetypal Analysis using $k = 1, \ldots, 25$ archetypes and 10 runs per k. Again, for each k select the fit with lowest RSS and plot the RSS as a function of k.

   b. Generate a 5×5 grid of images of the 25 archetypes found from the final iteration of your computation from part (b). You will need to expand these feature vectors to 784-dimensions maintaining the zero columns found in part 5a.

c. Based on your results from problems 5 and 6, compare the default *NMF* algorithm with the *archetypal* analysis.