Lecture 10 low rank models

wbg231

December 2022

introduction

- the prof stopped making videos so form now on these are going to be text book notes
- link
- up to now we have seen datasets where each data point is associated with a single entity
- now we are going to talk about cases where the data set are associated with two different entities
- we represent such data as entries in a matrix D for each enters $D[i, j]$ column i row j indicate the entries associated to each data point
- for example in recommender systems $D[i, j]$ would correspond to the rating user i gave to product j.
- for example in genomics $D[i, j]$ would correspond to the expression of gene i in cell j
- also i may use $D[i, j]$ and $D_{i.j}$ interchangeably
- in order to analyze such data it is often useful to represent each data point in terms of a small number of factors.
- this can be done by fitting the data using a low rank approximation

movie example

- consider a matrix of movie rating $D \in \mathbb{R}^{n_1 \times n_2}$ where $D_{i,j}$ is the rating user i gave to movie j
- to model D we approximate $D_{i,j}$ as a sum of r components where $r \leq minn_1, n_2$ (recall that the max number of eigenvalues the matrix $A \in \mathbb{R}^{n \times d}$ is min(n,d) that is the number of linearly independent rows and columns so if we use more than that number of components it is not a low rank model)

- let the lth component be the produce of a coefficient $a_l[i]$ associated to movie i and a coefficient $b_l[j]$ associated to user j thus we have approximation of the matrix given by

$$L[i,j] = \Sigma_{l=1}^{r} a_l[i] b_l[j], \quad \forall i \in [1, n_1], j \in [1, n_2]$$

- so keep in mind that what we have effectively is

$$L \in \mathbb{R}^{n_1 \times n_2} = \begin{pmatrix} a_1 & a_2 & \cdots & a_r \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \cdots \\ a_r \end{pmatrix} = AB \approx D \in \mathbb{R}^{n_1 \times n_2} : \forall a_i \in A a_i \in \mathbb{R}^{n_1}, b_i \in B b_i \in \mathbb{R}^{n_2}$$

- think of $a_l[i]$ as indicating wether the association of user i to factor l is positive negative or negligible.
- the model is bilinear because the approximation is a bilinear function of teh coefficients. if we fix either the user or movie coefficients the model is linear
- the rank of matrix L is equal to r. this can be seen as $rank(AB) = min(rank(A), (rank(b))) = min(min(n_1, r), min(min(r, n_2)) \leq r$ so if we select our component vectors to be linearly independent this will holds
- given that r is much smaller than the rank of D this is known as a low rank model.
- these models are fit using SVD which decomposes any matrix into rank 1 components

low rank models and PCA

- in a low rank model each column of $L[i] = A_i^T B_i$ that is a linear combination of basis vectors
- think of $D \in \mathbb{R}^{n_1 \times n_2} = \{D_1, \cdots D_{n_2}\}$ that is a set of n_1 dimensional data points
- from here we can reduce there dimensionality to r by using pca
  1. we can write each of the n_2 columns of D as $D[:, j] \in \mathbb{R}^{n_1}$
  2. first we can find the covariance matrix of columns of D as

  $$\Sigma_{cols} = \frac{1}{n_2} \Sigma_{j=1}^{n_2} D[:, j] D[:, j]^t = \frac{1}{n_2} DD^T$$

  3. take the eigenvectors associated with the top r largest eigenvalues tp obtain our principle directions

4. project the matrix onto the those r principle directions to get r principle components

$$w_l(j) = u_l^t D[:, j] \quad \forall j \in [1, n_2]$$

where u_i is the ith principle direction of Σ_{col} and

5. thus the set $\{w_1 \cdots w_r\}$ representing $D \in \mathbb{R}^{r \times n_2}$
6. and finally we can project back to our original space to get the low rank model $L_{pca-cols}[i,j] := \Sigma_{l=1}^{r} u_l[i] w_l[j] \forall i \in [1, n_1], j \in [1, n_2]$

- this model is optimal from the point of preserving mean squared ℓ_2 norm of the columns of D
- alternatively we could conduct the same operation on the rows of D in order to obtain the low rank model $L_{pca-rows}$
- we will now show that these approaches are equivalent

the singular value decomposition

- the SVD is a fundamental tool in linear algebra which decomposes a matrix into the product of a matrix with orthonormal columns a diagonal matrix and a matrix with orthonormal rows.
- **SVD** allows us to decomposes any matrix $A \in \mathbb{R}^{n_1 \times n_2}$ such that $n_1 \leq n_2$ (if not we can take the transpose and this will hold) with rank t to its singular value from as

$$A = \underbrace{\begin{bmatrix} u_1 & u_2 & \cdots & u_{n_1} \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} s_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \ddots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & s_t & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \ddots & \cdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix}}_{S} \underbrace{\begin{bmatrix} v_1 & v_2 & \cdots & v_{n_1} \end{bmatrix}^T}_{V^T},$$

where the singular values $s_1 \geq s_2 \geq \cdots \geq s_t$ are positive real numbers, the left singular vectors $u_1, u_2 \cdots u_r \in \mathbb{R}^{n_1}$ are orthonormal and the right singular vectors $v_1, v_2 \cdots v_r \in \mathbb{R}^{n_2}$

- proof
  1. first it is clear that for an arbitrary matrix $A \in \mathbb{R}^{n_1 \times n_2}$ we can write a symmetric matrix as $M := AA^T \in \mathbb{R}^{n_1 \times n_1}$
  2. and further by the spectral theorem we know that any symmetric matrix can be eiegen decomposed and thus has n_1 orthonormal vectors $u_1, \cdots, u_{n_1}$

3. further note that the corresponding eigenvalues λ_{n_1} are non-negative since $0 > ||A^t u_j||_2^2 = u_j^t A A^t u_j = u_j^t M u_j = \lambda_j u_j^t u_j = \lambda_j$ that is positive definite
4. and further the number of non-zero eigenvalues is equal to teh rank of A, and further because A and AA^T have he same rank
5. finish the proof top of page 466