
CSE 291. Assignment 2

Out: Tue Apr 19

Due: Tue May 3

2.1 Matrix factorizations

Download the data set for this problem from the course web site. The data set consists of grayscale images of handwritten digit TWOS. Some of these images are shown below.



The images are stored in MATLAB format as a matrix X with D rows and N columns, where $D = 784$ is the number of pixels per image and $N = 5958$ is the number of images. The i^{th} image in the data set can be displayed using the command:

```
imagesc(reshape(X(:, i), 28, 28));
```

In this problem, you will compute various low rank factorizations of the matrix $X \approx VY$, where V is a $D \times k$ matrix and Y is a $k \times N$ matrix, with $k = 25$. You will also explore the representations that these factorizations discover. Turn in your source code along with the results requested below.

(a) **Vector quantization**

Minimize the approximation error $\|X - VY\|$ subject to the constraints that $Y_{\alpha n} \in \{0, 1\}$ and $\sum_{\alpha} Y_{\alpha n} = 1$. Initialize the k columns of V using the first k columns of X . From your final solution, display the columns of V as images, and turn in a print-out of these images.

(b) **Principal component analysis**

Subtract out the mean image from each column of X , and call the resulting matrix $\bar{X}$. Minimize the approximation error $\|\bar{X} - VY\|$ subject to the constraint that the columns of V are orthonormal. From your final solution, display the columns of V as images, and turn in a print-out of these images. Also display the mean image.

(c) **Nonnegative matrix factorization**

Minimize the approximation error $\|X - VY\|$ subject to the constraint that matrices V and Y are nonnegative. Initialize the k columns of V using the first k columns of X , and initialize the matrix Y by setting every element equal to $\frac{1}{k}$. From your final solution, display the columns of V as images, and turn in a print-out of these images. (Note: some pixels have zero values across the whole set of images. You do not need to update the matrix elements in V corresponding to these pixels.)

2.2 Auxiliary function for logistic regression

In class we derived two bounds, one loose and one tight, on the sigmoid function $\sigma(\theta) = (1 + e^{-\theta})^{-1}$. In this problem, you will compare these bounds and justify the tighter one. Consider the functions:

$$\begin{aligned} f(\theta) &= -\log \sigma(\theta) \\ g(\theta, \theta_0) &= f(\theta_0) + f'(\theta_0)(\theta - \theta_0) + \frac{1}{8}(\theta - \theta_0)^2 \\ h(\theta, \theta_0) &= f(\theta_0) + f'(\theta_0)(\theta - \theta_0) + \left[\frac{\tanh(\theta_0/2)}{4\theta_0} \right] (\theta - \theta_0)^2 \end{aligned}$$

- (a) Plot the functions $f(\theta)$, $g(\theta, \theta_0)$, and $h(\theta, \theta_0)$ for $\theta_0 = 0$.
 - (b) Plot the functions $f(\theta)$, $g(\theta, \theta_0)$, and $h(\theta, \theta_0)$ for $\theta_0 = 2$.
 - (c) Plot the functions $f(\theta)$, $g(\theta, \theta_0)$, and $h(\theta, \theta_0)$ for $\theta_0 = -2$.
 - (d) Completing the proof in class, show that $\log(\cosh \theta)$ is concave in the variable θ^2 .
-

2.3 Convex quadratic programming with nonnegativity constraints

In this problem, you will derive an auxiliary function and iterative updates for convex quadratic programming with nonnegativity constraints. Your updates will generalize the multiplicative updates for nonnegative matrix factorization (Lee & Seung, 2001). *Hint:* the derivations in that paper may be useful for this problem. To begin, let $\mathbf{A} \in \mathbb{R}^{n \times n}$ denote a (symmetric) positive semidefinite matrix, and consider the quadratic form:

$$F(\mathbf{v}) = \frac{1}{2} \mathbf{v}^\top \mathbf{A} \mathbf{v} + \mathbf{b}^\top \mathbf{v}.$$

To derive an auxiliary function, let $\mathbf{A}^\pm \in \mathbb{R}^{n \times n}$ denote nonnegative matrices such that $\mathbf{A} = \mathbf{A}^+ - \mathbf{A}^-$. For example, we might take:

$$\begin{aligned} \mathbf{A}^+ &= \max(\mathbf{A}, 0), \\ \mathbf{A}^- &= \max(-\mathbf{A}, 0), \end{aligned}$$

where these operations are meant to be understood elementwise. In the above decomposition, $\mathbf{A}^+$ retains all the positive elements of $\mathbf{A}$, while $\mathbf{A}^-$ retains all the negative elements of $\mathbf{A}$.

- (a) For all nonnegative vectors $\mathbf{w}$, $\mathbf{v}$ and nonnegative matrices $\mathbf{A}^+$, prove the inequality:

$$\mathbf{v}^\top \mathbf{A}^+ \mathbf{v} \leq \sum_{i=1}^n \frac{(\mathbf{A}^+ \mathbf{w})_i}{w_i} v_i^2.$$

- (b) For all nonnegative vectors $\mathbf{w}$, $\mathbf{v}$ and nonnegative matrices $\mathbf{A}^-$, prove the inequality:

$$\mathbf{v}^\top \mathbf{A}^- \mathbf{v} \geq \sum_{i,j=1}^n \mathbf{A}_{ij}^- w_i w_j \left[1 + \log \frac{v_i v_j}{w_i w_j} \right].$$

(c) Consider the function:

$$G(\mathbf{v}, \mathbf{w}) = \frac{1}{2} \sum_{i=1}^n \frac{(\mathbf{A}^+ \mathbf{w})_i}{w_i} v_i^2 - \frac{1}{2} \sum_{i,j=1}^n \mathbf{A}_{ij}^- w_i w_j \left[1 + \log \frac{v_i v_j}{w_i w_j} \right] + \mathbf{b}^\top \mathbf{v}.$$

Show that this function satisfies the properties:

- (i) $G(\mathbf{v}, \mathbf{v}) = F(\mathbf{v})$,
- (ii) $G(\mathbf{v}, \mathbf{w}) \geq F(\mathbf{v})$.

Hence, conclude that it is an auxiliary function.

(d) Consider the iterative update rule $\mathbf{v}_{t+1} = \min_{\mathbf{v}} G(\mathbf{v}, \mathbf{v}_t)$. Show that this procedure leads to the elementwise multiplicative update:

$$v_i \leftarrow v_i \left[\frac{-b_i + \sqrt{b_i^2 + 4a_i c_i}}{2a_i} \right],$$

where $a_i = (\mathbf{A}^+ \mathbf{v})_i$ and $c_i = (\mathbf{A}^- \mathbf{v})_i$.

(e) Show that at fixed points of this update rule, either $v_i = 0$ or $\partial F / \partial v_i = 0$: that is, either v_i is pinned to the origin, or its value minimizes $F(\mathbf{v})$ given the other elements of $\mathbf{v}$ at the fixed point.

2.4 Project proposal

Submit a short, one-paragraph proposal for your final project. Projects can take many forms: e.g., a survey of 1-3 papers that follow up on material covered in lecture, an empirical comparison of competing algorithms, an interesting application of unsupervised learning, etc. Projects should relate to major themes of the course!

In the last two weeks of the quarter, students will give short presentations on their projects. Each presentation will be allotted 10-15 minutes. The most important criterion for projects is that these presentations teach something interesting to the other students in the class. A short written summary of each project (no longer than two pages) is also due at the end of the quarter.

If you have travel constraints during the last two weeks of the quarter, please let me know at this time.