Homework 6, CS 4220, Spring 2021
Instructor: Austin R. Benson
Due Friday, May 14, 2021 at 3:44pm ET on CMS (before lecture)

Policies

Submission. Submit your write-up as a single PDF on CMS: `https://cmsx.cs.cornell.edu`.
Coding questions. You can use any programming language for the coding parts of the assignment. Include your code in your write-up.
Typesetting. Your write-up must be typeset with LaTeX.
Collaboration. Please discuss and collaborate on the homework, but you have to write your own solutions and code.
Resources and attribution. Feel free to use any resources that might be helpful, and provide attribution for any key ideas. We only ask that you work on the problems in earnest. Please do not hunt for solutions with a search engine.

Problems

1. *Inequality quadratic.*

   Let $A \in \mathbb{R}^{n \times n}$ be symmetric and nonsingular, and consider the optimization problem

$$\begin{array}{ll} \underset{x \in \mathbb{R}^n}{\text{minimize}} & x^T A x \\ \text{subject to} & x^T x \leq 1. \end{array}$$

   (a) Write down the KKT conditions.

   (b) Characterize the global minimizer(s).

2. *ℓ_∞ regression*

   (a) Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a convex function. Show that $g\colon \mathbb{R}^d \to \mathbb{R}$ given by $g(x) = f(Mx + c)$ for $M \in \mathbb{R}^{n \times d}$ and vector $c \in \mathbb{R}^n$, is a convex function.

   (b) Show that $f\colon \mathbb{R}^n \to \mathbb{R}_{\geq 0}$, $f(x) = \|x\|$ is a convex function for any norm.

   (c) Let $A \in \mathbb{R}^{m \times n}$ with $m > n$, and let $b \in \mathbb{R}^m$. The function $f(x) = \|Ax - b\|_\infty$ is convex by parts (a) and (b), but it is not smooth. Reformulate

$$\underset{x}{\text{minimize}} \;\; \|Ax - b\|_\infty$$

   as a linear program (an optimization problem with a linear objective function and all linear constraints).

3. *Nonnegative Matrix Factorization.*

   Let $A \in \mathbb{R}^{m \times n}$ be a matrix with $A \geq 0$ entry-wise. The *nonnegative matrix factorization (NMF)* problem seeks to solve

$$\begin{array}{ll} \underset{W \in \mathbb{R}^{m \times k}, Z \in \mathbb{R}^{n \times k}}{\text{minimize}} & \|A - WZ^T\|_F^2 + \beta \|W\|_F^2 + \beta \|Z\|_F^2 \\ \text{subject to} & W \geq 0, \;\; Z \geq 0. \end{array}$$

Recall the alternating algorithm that we developed in Question 5a of Homework 2. If we had no constraints and fixed W, then we could find the optimal Z. Similarly, if we had no constraints and fixed Z, we could solve for the optimal W.

One possibility would then be to run a projection method, where we solve for the optimal W or Z (ignoring constraints) and then project onto the constraints by setting the negative entries to zero. A problem is that this may not decrease the objective function at every step, so we will look at an alternating *projected gradient descent* method instead.

(a) Let $f(W, Z) = \|A - WZ^T\|_F^2 + \beta\|W\|_F^2 + \beta\|Z\|_F^2$ be the objective function. Find the gradient of f with respect to W (when Z is fixed) and with respect to Z^T (when W is fixed).

(b) Implement an alternating projected gradient method for approximately solving the NMF problem. This means that in each alternating step, we first take a gradient descent step, which may violate the constraints. We then project onto the constraints; for this problem, this just means setting negative values in W or Z to zero.

Include a line search that attempts to find an updated factor that decreases the objective (after the projection). If your line search fails to find such a factor, provide a fallback strategy that might increase the objective.

For this part, you only need to turn in code.

(c) Download the Fashion-MNIST data, which consists of 70,000 28×28 images. In Julia, `MLDatasets.jl` is a useful package for this. See the README at https://github.com/zalandoresearch/fashion-mnist for other ways of downloading and reading the data.

Take the subset of images corresponding to T-Shirts, and use this data to construct a nonnegative matrix $A \in \mathbb{R}^{7{,}000 \times 784}$. Run your algorithm for 100 iterations with $k = 5$ and $\beta = 0.0001$. Show your computed Z factor by converting the 5 columns back into 28×28 images, and including these images in your write-up.