Math Foundations of ML, Fall 2017

Homework #6

Due Friday October 20, at the beginning of class

As stated in the syllabus, unauthorized use of previous semester course materials is strictly prohibited in this course.

1. Using you class notes, prepare a 1-2 paragraph summary of what we talked about in class in the last week. I do not want just a bulleted list of topics, I want you to use complete sentences and establish context (Why is what we have learned relevant? How does it connect with other things you have learned here or in other classes?). The more insight you give, the better.

2. Let $\{\boldsymbol{a}_1, \boldsymbol{a}_2, \ldots, \boldsymbol{a}_N\}$ be a set of linearly independent vectors in $\mathbb{R}^M$ (so clearly $N \leq M$). From these vectors, we generate another sequence of vectors $\{\boldsymbol{q}_1, \boldsymbol{q}_2, \ldots, \boldsymbol{q}_N\}$ using the algorithm[1]

$$\boldsymbol{q}_1 = \frac{\boldsymbol{a}_1}{\|\boldsymbol{a}_1\|_2}$$

then for $k = 2, 3, \ldots$,

$$\boldsymbol{w}_k = \boldsymbol{a}_k - \sum_{\ell=1}^{k-1} \langle \boldsymbol{a}_k, \boldsymbol{q}_\ell \rangle \boldsymbol{q}_\ell$$
$$\boldsymbol{q}_k = \frac{\boldsymbol{w}_k}{\|\boldsymbol{w}_k\|_2}.$$

   (a) Find $\boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{q}_3$ when

$$\boldsymbol{a}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \quad \boldsymbol{a}_2 = \begin{bmatrix} 1 \\ 1 \\ -1 \\ -1 \\ -1 \end{bmatrix}, \quad \boldsymbol{a}_3 = \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \\ 1 \end{bmatrix}$$

   (b) Show that $\{\boldsymbol{q}_1, \ldots, \boldsymbol{q}_N\}$ is an orthobasis for $\mathrm{Span}(\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_N\})$.

   (c) Explain, in a cogent manner, how the algorithm above can be interpreted as finding a $M \times N$ $\boldsymbol{Q}$ with orthonormal columns and an $N \times N$ upper triangular $\boldsymbol{T}$ such that $\boldsymbol{Q} = \boldsymbol{A}\boldsymbol{T}$.

   (d) Prove or disprove: an upper triangular matrix is invertible if and only none of the elements along the diagonal are zero. How do we know $\boldsymbol{T}$ from the previous part is invertible?

   (e) Explain how we can use the algorithm above to find the QR factorization $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$, where $\boldsymbol{R}$ is an upper triangular matrix.

[1] This is called the *Gram-Schmidt* algorithm.

(f) Suppose that an $M \times N$ matrix $\boldsymbol{A}$ has full column rank. Show that the solution to the least-squares problem

$$\underset{\boldsymbol{x}\in\mathbb{R}^N}{\text{minimize}} \ \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2$$

is $\hat{\boldsymbol{x}} = \boldsymbol{R}^{-1}\boldsymbol{Q}^{\mathrm{T}}\boldsymbol{y}$, where $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$ is the QR decomposition of $\boldsymbol{A}$.

3. Let $\boldsymbol{A}$ be a banded tri-diagonal matrix:

$$\boldsymbol{A} = \begin{bmatrix} d_1 & c_1 & 0 & 0 & 0 & \cdots & 0 \\ f_1 & d_2 & c_2 & 0 & 0 & \cdots & 0 \\ 0 & f_2 & d_3 & c_3 & 0 & \cdots & 0 \\ 0 & 0 & f_3 & d_4 & c_4 & \cdots & 0 \\ \vdots & \vdots & & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & & f_{N-2} & d_{N-1} & c_{N-1} \\ 0 & 0 & 0 & \cdots & & f_{N-1} & d_N \end{bmatrix}$$

(a) Argue that the LU factorization of $\boldsymbol{A}$ has the form

$$\boldsymbol{A} = \begin{bmatrix} * & 0 & 0 & 0 & \cdots & 0 \\ * & * & 0 & 0 & \cdots & 0 \\ 0 & * & * & 0 & \cdots & 0 \\ 0 & 0 & * & * & \cdots & 0 \\ \vdots & & & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & & * & * \end{bmatrix} \begin{bmatrix} * & * & 0 & 0 & 0 & \cdots & 0 \\ 0 & * & * & 0 & 0 & \cdots & 0 \\ 0 & 0 & * & * & 0 & \cdots & 0 \\ \vdots & & & & \ddots & \ddots & \\ 0 & 0 & \cdots & & & * & * \\ 0 & 0 & \cdots & & & 0 & * \end{bmatrix},$$

where $*$ signifies a non-zero term.

(b) Write down an algorithm that computes the LU factorization of $\boldsymbol{A}$, meaning the $\{\ell_i\}, \{u_i\}$, and $\{g_i\}$ below

$$\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ \ell_1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & \ell_2 & 1 & 0 & \cdots & 0 \\ 0 & 0 & \ell_3 & 1 & \cdots & 0 \\ \vdots & & & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & & \ell_{N-1} & 1 \end{bmatrix} \begin{bmatrix} u_1 & g_1 & 0 & 0 & 0 & \cdots & 0 \\ 0 & u_2 & g_2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & u_3 & g_3 & 0 & \cdots & 0 \\ \vdots & & & & \ddots & \ddots & \\ 0 & 0 & \cdots & & & u_{N-1} & g_{N-1} \\ 0 & 0 & \cdots & & & 0 & u_N \end{bmatrix},$$

I will get you started:

$$\begin{aligned} & u_1 = d_1 \\ & \text{for } k = 2, \ldots, N \\ & \qquad g_{k-1} = \\ & \qquad \ell_{k-1} = \\ & \qquad u_k = \\ & \text{end} \end{aligned}$$

(c) What is the computational complexity of the algorithm above? (How does the number of computations scale with N?) Once the LU factorization is in hand, how does solving $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$ scale with N?

4. A function $f : \mathbb{R} \to \mathbb{R}$ of one variable that is twice differentiable is convex if its second derivative is positive everywhere. A function $f : \mathbb{R}^M \to \mathbb{R}$ of M variables is convex if every function of one variable formed by looking along a ray starting from an arbitrary point has a second derivative that is positive anywhere. That is, the function

$$g_{\boldsymbol{x},\boldsymbol{v}}(t) = f(\boldsymbol{x} + t\boldsymbol{v})$$

is a convex function of one variable (t) for all $\boldsymbol{x}, \boldsymbol{v}$.

Let $\boldsymbol{H}$ be a symmetric matrix. Show that

$$f(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{T}} \boldsymbol{H} \boldsymbol{x}$$

is convex if and only if $\boldsymbol{H}$ is symmetric positive semi-definite. That is, all the eigenvalues of $\boldsymbol{H}$ are non-negative, or equivalently $\boldsymbol{w}^{\mathrm{T}} \boldsymbol{H} \boldsymbol{w} \geq 0$ for all $\boldsymbol{w} \in \mathbb{R}^M$.