ECE 3803, Fall 2021

Homework #2

Due Tuesday, September 14, at 11:59pm

1. Prepare a one 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 classes?). The more insight you give, the better.

2. Compute the gradient and Hessian matrix of the following functions. Note that $\boldsymbol{x}$ is a vector in $\mathbb{R}^N$ in all the problems below.
   (a) $f(\boldsymbol{x}) = \boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}^T\boldsymbol{x} + c$, where $\boldsymbol{A}$ is an $N \times N$ symmetric matrix (i.e., $\boldsymbol{A} = \boldsymbol{A}^T$),$\boldsymbol{b}$ is an $N \times 1$ vector, and c is a scalar.
   (b) $f(\boldsymbol{x}) = -\cos(2\pi\boldsymbol{x}^T\boldsymbol{x}) + \boldsymbol{x}^T\boldsymbol{x}$.
   (c) $f(\boldsymbol{x}) = \sum_{m=1}^{M} \log(1 + e^{-\boldsymbol{a}_m^T\boldsymbol{x}})$, where $\boldsymbol{a}_1, \dots, \boldsymbol{a}_M$ are $N \times 1$ vectors.

3. Using the properties of an inner product, prove that the Pythagorean Theorem holds for any "induced norm". Specifically, show that if $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$, then for any $\boldsymbol{x}, \boldsymbol{y}$ satisfying $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$, $\|\boldsymbol{x} + \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2$.

4. Using the properties of an inner product, prove that the Cauchy-Schwarz inequality holds for any "induced norm".
   (a) Specifically, show that if $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$, then for any $\boldsymbol{x}, \boldsymbol{y}$ we have $|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leq \|\boldsymbol{x}\|\|\boldsymbol{y}\|$. [Hint: Consider writing $\boldsymbol{x}$ as a linear combination of $\boldsymbol{y}$ and the vector $\boldsymbol{z} = \boldsymbol{x} - \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{y}\|^2}\boldsymbol{y}$. What is $\langle \boldsymbol{z}, \boldsymbol{y} \rangle$? What can we infer from the previous problem?]
   (b) *Optional: Show that equality holds if and only if $\boldsymbol{y} = a\boldsymbol{x}$ for some $a \in \mathbb{R}$.

5. Prove that if $\boldsymbol{A}$ and $\boldsymbol{B}$ are square $N \times N$ matrices, then if $\boldsymbol{A}\boldsymbol{B} = \mathbf{I}$, we must also have $\boldsymbol{B}\boldsymbol{A} = \mathbf{I}$. Some hints to help you get started:
   - An equivalent statement to $\boldsymbol{B}\boldsymbol{A} = \mathbf{I}$ is that, for any $\boldsymbol{x} \in \mathbb{R}^N$, $\boldsymbol{B}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}$.
   - Think about what can you say about $\mathcal{R}(\boldsymbol{A}\boldsymbol{B})$.
   - Think about what you can say about the relationship between $\mathcal{R}(\boldsymbol{A}\boldsymbol{B})$ and $\mathcal{R}(\boldsymbol{B})$.
   - Recall that for any $\boldsymbol{x} \in \mathcal{R}(\boldsymbol{B})$, we can write $\boldsymbol{x}$ as a linear combination of the columns of $\boldsymbol{B}$, i.e., $\boldsymbol{x} = \boldsymbol{B}\boldsymbol{a}$ for some vector $\boldsymbol{a} \in \mathbb{R}^N$.

6. The standard least squares objective is to minimize

$$\|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y}\|_2^2 = \sum_{m=1}^{M} (\boldsymbol{a}_m^T\boldsymbol{x} - y_m)^2,$$

where $\boldsymbol{a}_m^T$ represents the m^{th} row of $\boldsymbol{A}$. This inherently assumes that finding an explanation for each y_m is equally important. There are often situations where you might have more confidence in some measurements over others. For instance, you may have different sensors which yield measurements that have been corrupted by different levels of noise. If you know how much noise has been added to each y_m, you can compensate for this by introducing weights w_m and consider the modified objective function

$$\sum_{m=1}^{M} (\boldsymbol{a}_m^T \boldsymbol{x} - y_m)^2 = \sum_{m=1}^{M} w_m (\boldsymbol{a}_m^T \boldsymbol{x} - y_m)^2. \tag{1}$$

One could set the weights w_m to be higher for the y_m which are known to be reliable, and lower for the y_m that one suspects are particularly noisy.

(a) Show that the so-called *weighted least squares* objective function in (1) can be written as

$$\|\boldsymbol{W}(\boldsymbol{A}\boldsymbol{x} - \boldsymbol{y})\|_2^2$$

for a particular choice of $\boldsymbol{W}$.

(b) Derive a system of equations that the $\boldsymbol{x}$ that minimizes (1) must satisfy by taking the gradient with respect to $\boldsymbol{x}$ and setting this to zero (see page 17 of the notes for the analogous derivation in the unweighted case).

7. Suppose that we are given a set of points $(x_1, y_1), \ldots, (x_M, y_M)$ and wish to interpolate between these points using a polynomial of degree $N-1$, i.e., we would like to find a function of the form

$$p(x) = a_{N-1}x^{N-1} + \cdots + a_1 x + a_0,$$

such that $p(x_m) = y_m$ for $m = 1, \ldots, M$.

(a) The problem of finding such an interpolating polynomial can be cast as finding the solution to a system of equations of the form $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{a}$, where $\boldsymbol{y}$ is an $M \times 1$ matrix and $\boldsymbol{X}$ is an $M \times N$ matrix, both of which are determined by our observations, and $\boldsymbol{a}$ is an $N \times 1$ vector that represents the parameters of the polynomial. Write down what the entries of $\boldsymbol{y}$, $\boldsymbol{X}$ and $\boldsymbol{a}$ are.

(b) It is a fact that if you have M samples of a function, you can always perfectly interpolate those M samples using a polynomial of degree $M-1$ (i.e., a polynomial with M parameters). Given a function $f(x)$, write a Python script that will

- Form a dataset of size M by computing M equally spaced samples $x_1, \ldots, x_M$ in the interval $[-1, 1]$ and setting $y_m = f(x_m)$.
- Form the system $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{a}$.
- Solve for $\boldsymbol{a}$ by inverting $\boldsymbol{X}$.
- Plot the original $f(x)$ along with your interpolating polynomial.

Test your code on the function $f(x) = 1 - x^2$ (verifying that with $M \geq 3$ your interpolation identifies $f(x)$ perfectly.)

(c) Now test your code on the function

$$f(x) = \frac{1}{1 + 25x^2}.$$

Note that this is *not* a polynomial, but it is smooth and can be well-approximated by a polynomial. One way to find a polynomial that might be a good approximation would be to sample the function and then interpolate. Do this for $M = 3, 5, 7, \ldots, 21$. (Since this is an even function, we expect the polynomial to only have non-zero coefficients for terms involving the even powers of x.) Decide which value of M yields the best result.[1] Submit plots showing $f(x)$ and the interpolating polynomials for both the best choice of M as well as for $M = 21$.

(d) Up to now we have focused only on the interpolation problem, in which we use M samples to fit a polynomial of degree $M - 1$. Now suppose we wish to fit a polynomial of degree $N - 1$ where $N < M$. Argue that we can cast this as a least squares optimization problem – write down both the optimization problem and an analytical formula for the solution. Modify your code from part (b) to handle the general case where $N \leq M$ (you should only need to make minor modifications outside of one line of code which will need to be replaced) and use this to find a polynomial of degree 20 that approximates the function from part (c) by setting $M > 20$. You should experiment with M and can set it as large as you like. Comment on your results and compare to what you obtained in part (c).

(e) *Optional: Implement the weighted least squares approach developed in Problem 6 and experiment with different "weighting schemes" in the setting of part (d). You might want to try placing comparatively less weight on the samples near the boundary. By doing this can you increase M and achieve better results than before?

[1] To do this automatically, I would use some kind of numerical integration here to estimate $\int_{-1}^{1}(f(x) - p_M(x))^2 dx$ where $p_M(x)$ is the interpolating polynomial of degree $M - 1$. This can be a very simple method. Recall the Riemann approximation of an integral.