Fall 2022 *Instructor: Ashwin Pananjady*

7750: Mathematical Foundations of Machine Learning

Linear algebra and probability for data analysis

Homework 1

Released: Aug 24 *Due: Sep 6, 11:59pm ET*

Note. All external sources and collaborators must be acknowledged in your submission. As stated in the syllabus, unauthorized use of previous semester course materials is strictly prohibited in this course.

Objective. To solidify your algebraic and geometric understanding of linear representations with basis functions. To build intuition for normed vector spaces by reviewing some finite-dimensional linear algebra.

Resources. Lectures, notes, and modules posted before Aug 31.

Notation: Capital boldface letters will be matrices, and small boldface letter will be vectors. Capital letters (not boldface) will be random variables (and sometimes random vectors), and small letters will typically be scalars. Dimensions of matrices and vectors will be specified when needed, but for the most part, you should be able to intuit these yourself (and doing this is a useful exercise).

Problem 1 (Visualizing quadratic functions and linear algebra review). 20 points: For $\mathbf{x} \in \mathbb{R}^d$, consider the function $f(\mathbf{x}) = \mathbf{x}^\top \mathbf{A}\mathbf{x} + \mathbf{b}^\top \mathbf{x} + c$ for a square and symmetric matrix $\mathbf{A} \in \mathbb{R}^{d \times d}$, vector $\mathbf{b} \in \mathbb{R}^d$ and scalar c.

(a) Write the gradient of f. Suppose the matrix $\mathbf{A}$ is invertible. Is there a unique solution to $\nabla f(\mathbf{x}) = \mathbf{0}$, and if so, what is it?

(b) You have been given an iPython notebook with starter code to generate 3D plots and contour plots of functions. Fill in the requisite lines of code and provide both 3D and contour plots of f for the following settings of $(\mathbf{A}, \mathbf{b}, c)$:

- $$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad \mathbf{b} = \mathbf{0}, \quad c = 0.$$
- $$\mathbf{A} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}, \quad \mathbf{b} = \mathbf{0}, \quad c = 0.$$
- $$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}, \quad \mathbf{b} = \mathbf{0}, \quad c = 0.$$
- $$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad \mathbf{b} = \mathbf{0}, \quad c = 0.$$

- $$\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}, \quad \mathbf{b} = \mathbf{0}, \quad c = 0.$$

Hopefully this gives you some idea of the geometry of these functions. Comment on the contour plots that you observe in all cases and how they relate to the respective 3D plots.

Note: If you prefer to use your own code to generate these plots instead of the provided starter code, that is fine.

(c) Show that we have the linear representation $f(\mathbf{x}) = \sum_{j=1}^{M} \alpha_j p_j(\mathbf{x})$ where $p_1, \ldots, p_M$ are all the monomials in $(x_1, \ldots, x_d)$ of degree at most 2. In particular, for each such monomial write down the coefficient that multiplies it.

Takeaway: This and the previous part should convince you that there are quite a few "shapes" taken by functions that can be represented as quadratic polynomials in two variables.

(d) For the fifth example (non-diagonal $\mathbf{A}$), use a package to compute its eigendecomposition. Compare the orthogonal matrix returned by this eigendecomposition to what you see on the corresponding contour plot. Is $\mathbf{A}$ positive semidefinite (PSD)?

(e) Now return to the original problem in dimension d with general $(\mathbf{A}, \mathbf{b}, c)$. If $\mathbf{A}$ is PSD and invertible, i.e., positive definite[1], argue that $\widehat{\mathbf{x}} = -\frac{1}{2}\mathbf{A}^{-1}\mathbf{b}$ is the unique global minimum of f.

(Hint: There are several ways to do this, but one way is to write the function as $f(\mathbf{x}) = (\mathbf{x} - \widehat{\mathbf{x}})^\top \mathbf{A}(\mathbf{x} - \widehat{\mathbf{x}}) + \text{other terms}$ and argue from there.)

Problem 2 (Monomial basis functions and invertibility). 20 points:

(a) In class, we claimed that given a dataset $(x_i, y_i)_{i=1}^n$ where $x_i, y_i \in \mathbb{R}$ were distinct across i, there is a *unique* polynomial of degree $n-1$ that interpolates these points. Without proving uniqueness, we showed that one such polynomial interpolator of degree $n-1$ was given by the Lagrange polynomial

$$p(x) = \sum_{k=1}^{n} y_k \cdot \prod_{\substack{1 \le j \le n \\ j \ne k}} \frac{x - x_j}{x_k - x_j}.$$

In parts (a)-(e), you will prove the *uniqueness* claim for every finite n, provided $d = 1$.

Construct the square matrix

$$\mathbf{\Phi} = \begin{bmatrix} 1 & x_1 & x_1^2 & \ldots & x_1^{n-1} \\ 1 & x_2 & x_2^2 & \ldots & x_2^{n-1} \\ \vdots & & & & \vdots \\ 1 & x_n & x_n^2 & \ldots & x_n^{n-1} \end{bmatrix}.$$

Argue formally that the Lagrange polynomial $p(x)$ is the *unique* polynomial of degree $n-1$ that interpolates the points $\{x_i, y_i\}_{i=1}^n$ *if and only if* $\mathbf{\Phi}$ is invertible.

[1] A positive definite matrix satisfies $\mathbf{v}^\top \mathbf{A}\mathbf{v} \ge 0$ for all $\mathbf{v} \in \mathbb{R}^d$, with $\mathbf{v}^\top \mathbf{A}\mathbf{v} = 0$ if and only if $\mathbf{v} = \mathbf{0}$

(b) Next, show that if $n = 2$ and $x_1 \neq x_2$, then there is a unique interpolating polynomial of degree 1 that interpolates these points.

(c) We will now try to set up a general way to attack the invertibility of $\mathbf{\Phi}$ by calculating its determinant. In particular, we will form another matrix $\mathbf{\Phi}'$. Let M_j denote the j-th column of a matrix $\mathbf{M}$ and execute the following loop:

for each $j = 0, \dots n-1$: Set $\Phi'_{n-j} = \Phi_{n-j} - x_1 \cdot \Phi_{n-j-1}$.

Use the convention that $\Phi_0 = 0$. Argue that $\det(\mathbf{\Phi}) = \det(\mathbf{\Phi}')$.

(d) Show that $\mathsf{det}(\mathbf{\Phi}) = (x_2 - x_1) \times (x_3 - x_1) \times \cdots \times (x_n - x_1) \times \mathsf{det}(\overline{\mathbf{\Phi}})$, where we have defined the $(n-1) \times (n-1)$ matrix

$$\overline{\mathbf{\Phi}} = \begin{bmatrix} 1 & x_2 & x_2^2 & \dots & x_2^{n-2} \\ \vdots & & & & \vdots \\ 1 & x_n & x_n^2 & \dots & x_n^{n-2} \end{bmatrix}.$$

(e) Proceed iteratively and produce an expression for $\mathsf{det}(\mathbf{\Phi})$. Conclude that $\mathbf{\Phi}$ is invertible if and only if the points $\{x_i\}_{i=1}^n$ are distinct.

Note: The situation when $\mathbf{x}_i \in \mathbb{R}^d$ is significantly more delicate; indeed, uniqueness of the points does not suffice. However, if the points are sufficiently generic (say $\{\mathbf{x}_i\}_{i=1}^n$ are chosen i.i.d. from a standard Gaussian distribution $\mathsf{N}(0, \mathbf{I})$) then it can be shown that $\mathbf{\Phi}$ is indeed invertible. Bonus credit for proving this, but note that it is a challenging problem.

(f) The remaining parts are unrelated to the previous parts. We will now argue that the number of monomials of degree less than or equal to ℓ in d variables can be explicitly calculated. First argue that the number of such monomials can be counted by placing d "stars" and $\ell - 1$ "bars" in a sequence and counting the number of unique configurations that result when these are permuted.

(g) How many unique configurations are there above? Your expression should be explicit, and a function of the pair (d, ℓ).

Problem 3 (Splines and linear equations). 20 points: Suppose that $f : \mathbb{R} \to \mathbb{R}$ is a second-order spline defined by the overlap of five B-splines:

$$f(x) = \sum_{k=0}^{4} \alpha_k b_2(x - k),$$

where $b_2(x)$ is defined as on page 11 of the Fall 20 notes:

$$b_2(x) = \begin{cases} (x + 3/2)^2/2 & -3/2 \leq x \leq -1/2 \\ -x^2 + 3/4 & -1/2 \leq x \leq 1/2 \\ (x - 3/2)^2/2 & 1/2 \leq x \leq 3/2 \\ 0 & |x| \geq 3/2 \end{cases}.$$

(a) Write a function which takes $\alpha = \{\alpha_0, \ldots, \alpha_4\}$ and `x` as input and returns samples of $f(x)$ at the locations specified in the vector `x`. Turn in a plot of $f(x)$ for $\alpha = \{2, 1, -1, 3, -1\}$. Sample x densely enough (by specifying sufficiently many points in `x`) so that your plot looks like a smooth function.

(b) Suppose

$$f(0) = 2, \quad f(1) = 2, \quad f(2) = -5, \quad f(3) = -5, \quad f(4) = -2.$$

What are the corresponding α_k? (Hint: you will have to construct a system of equations then solve it. You don't necessarily have to do this by hand.)

(c) To generalize this, suppose that f is now a superposition of N B-splines, with

$$f(x) = \sum_{n=0}^{N-1} \alpha_n b_2(x - n).$$

Describe how to construct the $N \times N$ matrix that maps the coefficients α to the N samples $f(0), \ldots, f(N-1)$. That is, find a matrix $\mathbf{A}$ such that

$$\begin{bmatrix} f(0) \\ f(1) \\ \vdots \\ f(N-1) \end{bmatrix} = \mathbf{A} \cdot \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_{N-1} \end{bmatrix}$$

(d) Argue that the matrix $\mathbf{A}$ from part (c) is invertible for all values of N.
(Hint: Observe that the matrix $\mathbf{A}$ is banded and has large values on the diagonal. How might you use this to argue invertibility?)

(e) To take this even further, suppose that

$$f(x) = \sum_{n=-\infty}^{\infty} \alpha_n b_2(x - n),$$

so f is described by the (possibly infinite) sequence of numbers $\{\alpha_n\}_{n \in \mathbb{Z}}$. Show that there is a convolution operator that maps the sequence $\{\alpha_n\}$ to the sequence $\{f(n)\}$. That is, find a sequence of numbers $\{h_n\}_{n \in \mathbb{Z}}$ such that

$$f(n) = \sum_{\ell=-\infty}^{\infty} h_\ell \, \alpha_{n-\ell}.$$

Problem 4. (Fun with norms and inner products). 20 points:

(a) For parts (a-e), $\langle \cdot, \cdot \rangle$ is the standard inner product on $\mathbb{R}^N$. Prove that $|\langle \mathbf{x}, \mathbf{y} \rangle| \le \|\mathbf{x}\|_\infty \cdot \|\mathbf{y}\|_1$.

(b) Prove that $\|\mathbf{x}\|_1 \le \sqrt{N} \cdot \|\mathbf{x}\|_2$. (Hint: Cauchy–Schwarz)

(c) Let B_2 be the unit ball for the ℓ_2 norm in $\mathbb{R}^N$. Show that

$$\max_{\mathbf{x} \in B_2} \langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{y}\|_2.$$

Describe the vector $\mathbf{x}$ (as a function of $\mathbf{y}$) which achieves the maximum.

(d) Let B_∞ be the unit ball for the ℓ_∞ norm in $\mathbb{R}^N$. Show that

$$\max_{\mathbf{x}\in B_\infty} \langle \mathbf{x}, \mathbf{y}\rangle = \|\mathbf{y}\|_1.$$

Describe the vector $\mathbf{x}$ (as a function of $\mathbf{y}$) which achieves the maximum.

(e) Let B_1 be the unit ball for the ℓ_1 norm in $\mathbb{R}^N$. Show that

$$\max_{\mathbf{x}\in B_1} \langle \mathbf{x}, \mathbf{y}\rangle = \|\mathbf{y}\|_\infty.$$

Describe the vector $\mathbf{x}$ (as a function of $\mathbf{y}$) which achieves the maximum.

(f) For parts (f-i), suppose you are given an $N \times N$ matrix $\mathbf{Q}$, and set

$$\langle \mathbf{x}, \mathbf{y}\rangle_Q = \mathbf{y}^\top \mathbf{Q}\mathbf{x},$$

for vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^N$. Prove that if $\mathbf{Q}$ has an entry along its diagonal that is nonpositive, then $\langle \cdot, \cdot \rangle_Q$ cannot be a valid inner product on $\mathbb{R}^N$.

(g) Prove that if $\mathbf{Q}$ is not symmetric, then $\langle \cdot, \cdot \rangle_Q$ cannot be valid inner product on $\mathbb{R}^N$.

(h) Recall that $\mathbf{Q}$ is *symmetric positive definite* if it is symmetric and obeys

$$\mathbf{x}^\top \mathbf{Q}\mathbf{x} > 0, \quad \text{for all } \ \mathbf{x} \in \mathbb{R}^N, \ \mathbf{x} \neq 0.$$

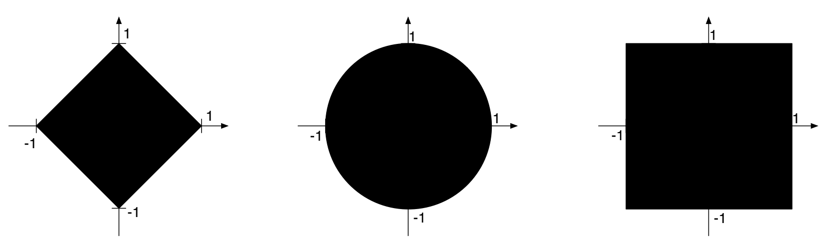
Prove that $\langle \cdot, \cdot \rangle_Q$ is a valid inner product on $\mathbb{R}^N$ if and only if $\mathbf{Q}$ is symmetric positive definite.

(i) Define the norm on $\mathbb{R}^2$

$$\|\mathbf{x}\| = \|\mathbf{A}\mathbf{x}\|_2, \quad \mathbf{A} = \begin{bmatrix} 3 & 3 \\ -1/2 & 1/2 \end{bmatrix}.$$

Find $\mathbf{Q}$ so that $\langle \cdot, \cdot \rangle_Q$ induces this norm.

Problem 5. (Visualizing norm balls). 20 points One way to visualize a norm in $\mathbb{R}^2$ is by its *unit ball*, the set of all vectors such that $\|\mathbf{x}\| \leq 1$. For example, we have seen that the unit balls for the ℓ_1, ℓ_2, and ℓ_∞ norms look like:



$B_1 = \{\mathbf{x} : \|\mathbf{x}\|_1 \leq 1\}$ $\quad B_2 = \{\mathbf{x} : \|\mathbf{x}\|_2 \leq 1\}$ $\quad B_\infty = \{\mathbf{x} : \|\mathbf{x}\|_\infty \leq 1\}$

Given an appropriate subset of the plane, $B \subset \mathbb{R}^2$, it might be possible to define a corresponding norm using

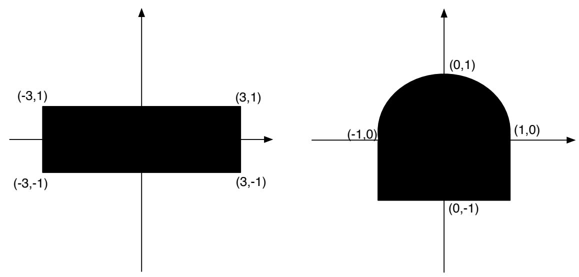
$$\|\mathbf{x}\|_B = \text{minimum value } r \geq 0 \text{ such that } \mathbf{x} \in rB, \tag{1}$$

where rB is just a scaling of the set B:

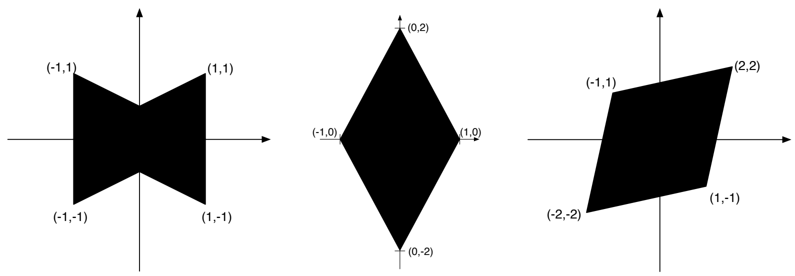
$$\mathbf{x} \in B \quad \Rightarrow \quad r \cdot \mathbf{x} \in rB.$$

(a) Let $\mathbf{x} = \begin{bmatrix} -4 \\ 2 \end{bmatrix}$. For $p = 1, 2, \infty$, find $r = \|\mathbf{x}\|_p$, and sketch $\mathbf{x}$ and rB_p (use different axes for each of the three values of p).

(b) Consider the 5 shapes below.



B_a B_b



B_c B_d B_e

Explain why $\|\cdot\|_{B_b}$ and $\|\cdot\|_{B_c}$ are **not** valid norms. The most convincing way to do this is to find vectors for which one of the three properties of a valid norm are violated.

(c) Give a concrete method for computing $\|\mathbf{x}\|_{B_a}$, $\|\mathbf{x}\|_{B_d}$, and $\|\mathbf{x}\|_{B_e}$ for any given vector $\mathbf{x}$. (For example: for B_1, which corresponds to the ℓ_1 norm, we would write $\|\mathbf{x}\|_1 = |x_1| + |x_2|$.) Using you expressions, show that these are indeed valid norms.

Problem 6 (Bonus). 5 points: What do you expect to learn from this class? Please be honest and as detailed as you would like. Note that while the theme of the class is theoretical and we will stick to it, but we may adjust our coverage slightly if there is enough interest to learn a particular topic. Are you struggling with any particular concept so far and do you have any suggestions for the course staff?