MSSC 5931 - Homework 1

Henri Medeiros Dos Reis

September 16, 2022

1. **Representing linear functions as matrix multiplication.** Suppose that $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is linear. Show that there is a matrix $A \in \mathbb{R}^{m\times n}$ such that for all $x \in \mathbb{R}^n$, $f(x) = Ax$. (Explicitly describe how you get the coefficients A_{ij} from f, and then verify that $f(x) = A$ for any $x \in \mathbb{R}^n$.) Is the matrix A that represents f unique? In other words, if $\tilde{A} \in \mathbb{R}^{mn}$ is another matrix such that $f(x) = \tilde{A}x$ for all $x \in \mathbb{R}^n$, then do we have $\tilde{A} = A$? Either show that this is so, or give an explicit counterexample.

   Solution 1. *Proof.*
$$X = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \in \mathbb{R}^n$$
   $x = x_1e_1 + x_2e_2 + \cdots + x_ne_n$, where e_i is the i^{th} unit vector from the identity matrix. Then
$$Ax = f(x) = \begin{bmatrix} f_1(e_1) & f_2(e_2) & \cdots & f_n(e_n) \end{bmatrix} x$$
   Finally, we can say that $A = \begin{bmatrix} f_1(e_1) & f_2(e_2) & \cdots & f_n(e_n) \end{bmatrix}$, where $f(e_i) \in \mathbb{R}^m$ If $\tilde{A}x = f(x) = Ax$ for all $x \in \mathbb{R}^n$, then $\tilde{A}$ has to be the same as A. Since $\tilde{A}x = Ax \Rightarrow \tilde{A}x - Ax = 0$
   Let x be e_i then $\tilde{A}e_i - A_ei = 0$, in the i^{th} column $\tilde{A} - A = 0$, then $\tilde{A} = A$. Therefore, A is unique $\square$

2. **Matrix representation of polynomial differentiation.** We can represent a polynomial of degree less than n,
$$p(x) = a_{n-1}x_{n-1} + a_{n-2}x_{n-2} + \cdots + a_1x + a_0,$$
   as the vector $(a_0, a_1, ..., a_{n-1}) \in \mathbb{R}^n$. Consider the linear transformation $\mathcal{D}$ that differentiates polynomials, i.e., $\mathcal{D}p = dp/dx$. Find the matrix D that represents $\mathcal{D}$ (i.e., if the coefficients of p are given by a, then the coefficients of dp/dx are given by Da).

   Solution 2. Since $\mathcal{D}$ is just a linear transformation, then $\mathcal{D} = \begin{bmatrix} \mathcal{D}\vec{v_1} & \mathcal{D}\vec{v_2} & \cdots & \mathcal{D}\vec{v_n} \end{bmatrix}$
   Let $\vec{v_1}$ be the vector that represents all constant numbers. $p_1(x) = C \Rightarrow \frac{dp_1}{dx}p_1(x) = 0$
$$\mathcal{D}\vec{v_1} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = 0$$

Let $\vec{v_2}$ be the vector that represents all x to the power of 1. $p_2(x) = x \Rightarrow \frac{dp_2}{dx} p_2(x) = 1$

$$\mathcal{D}\vec{v_2} = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = e_1$$

Also let $\vec{v_3}$ be the vector that represents all x to the power of 2. $p_3(x) = x^2 \Rightarrow \frac{dp_2}{dx} p_2(x) = 2x$

$$\mathcal{D}\vec{v_3} = \begin{bmatrix} 0 \\ 2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = 2e_2$$

Now by the pattern created let $\vec{v_n}$ be the vector that represents all x to the power of $n-1$. $p_n(x) = x^{n-1} \Rightarrow \frac{dp_n}{dx} p_n(x) = (n-1)x^{n-2}$

$$\mathcal{D}\vec{v_n} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ n-1 \\ 0 \end{bmatrix} = (n-1)e_{n-1}$$

Therefore the D matrix is

$$\begin{bmatrix} 0 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 2 & \cdots & 0 \\ \vdots & \vdots & 0 & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & \vdots & \vdots & \ddots & n-1 \\ 0 & 0 & 0 & \cdots & 0 \end{bmatrix}$$

3. **Counting paths in an undirected graph.** Consider an undirected graph with n nodes, and no self loops (i.e., all branches connect two different nodes). Let $A \in \mathbb{R}^{n \times n}$ be the node adjacency matrix, defined as

$$A_{ij} = \begin{matrix} 1 & \text{if there is a branch from node i to node j} \\ 0 & \text{if there is no branch from node i to node j} \end{matrix}$$

Note that $A = AT$, and $A_{ii} = 0$ since there are no self loops. We can interpret A_{ij} (which is either zero or one) as the number of branches that connect node i to node j. Let $B = A^k$, where $k \in Z, k \geq 1$. Give a simple interpretation of B_{ij} in terms of the original graph. (You might need to use the concept of a path of length m from node p to node q.)

Solution 3. $B = A^k$, where $k = 1, 2, 3, \cdots, n$
$B = A^1$, B_{ij} is just the number of connections between node i and node when its length is 1

$B = A^2 = AA$, B_{ij} is again the number of connections between node i and j, but now the length is going to be 2, because if you have an entry of 1 in both As then $A_{im}A_{mj}$ is of length 2 and $B_{ij} = \sum_{m=1}^{n} A_{im}A_{mj}$, gives you all the possible paths of length 2 in the graph.

Similarly when we consider $B = A^3 = AAA$, we will end up with all possible paths of length 3, since if any of A_1, A_2 or A_3 have 1s means a path of length 3.
Then $B_{ij} = \sum_{m_1=1}^{n} \sum_{m_2=1}^{n} A_{im_1}A_{m_1m_2}A_{m_2j}$

Finally $B = A^k$. Where B_{ij} will give the number of paths of length k.

$$B_{ij} = \sum_{m_1=1}^{n} \sum_{m_{k-1}=1}^{n} A_{im_1}A_{m_1m_2} \cdots A_{m_{k-1}j}$$

Where all entries $A_{im_1}, A_{m_1m_2}, \cdots, Am_{k-1}k$ are either 0 or 1, which means that when they are all 1s, there is a path of length k between nodes i and j

4. **Gradient of some common functions.** Recall that the gradient of a differentiable function $f\colon \mathbb{R}^n \to \mathbb{R}$, at a point $x \in \mathbb{R}^n$, is defined as the vector

$$\nabla f(x) = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{bmatrix}$$

where the partial derivatives are evaluated at the point x. The first order Taylor approximation of f, near x, is given by

$$\hat{f}_{tay}(z) = f(x) + \nabla f(x)^T(z - x).$$

This function is affine, i.e., a linear function plus a constant. For z near x, the Taylor approximation $\hat{f}_{tay}$ is very near f. Find the gradient of the following functions. Express the gradients using matrix notation.

(a) $f(x) = a^T x + b,$ where $a \in \mathbb{R}^n, b \in \mathbb{R}$.

(b) $f(x) = x^T A x,$ for $A \in \mathbb{R}^{n\times n}$.

(c) $f(x) = x^T A x,$ where $A = A^T \in \mathbb{R}^{n\times n}$. (Yes, this is a special case of the previous one.)

Solution 4. (a) $f(x) = (a^t)x + b$, where $A \in \mathbb{R}^n, b \in \mathbb{R}$ so when we expand a^t, then

$$a^t = \begin{bmatrix} a_1^t & a_2^t & \cdots & a_n^t \end{bmatrix}$$

Therefore $f(x) = a^t x + b = \begin{bmatrix} a_1^t & a_2^t & \cdots & a_n^t \end{bmatrix} x + b = (\sum_{i=1}^n a_i x_i) + b$

Then:

$$\nabla f(x) = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{bmatrix} = \frac{\partial f}{\partial x_1} a_1 x_1, \frac{\partial f}{\partial x_2} a_2 x_2, \cdots, \frac{\partial f}{\partial x_n} a_n x_n = a_1, a_2, \cdots, a_n \Rightarrow a_i$$

Therefore, the gradient of this function is a constant vector a

(b) Following the same path as previously; Let's expand $f(x) = x^t A x$, since x^t is a vector that multiplies Ax. We can write

$$f(x) = \sum_{i=1}^n x^t (Ax) i = x_1^t (Ax)_1 + x_2^t (Ax)_2 + \cdots + x_n^t (Ax)_n$$

Where Ax is also a linear combination of the matrix A, then

$$x_1^t (Ax)_1 + x_2^t (Ax)_2 + \cdots + x_n^t (Ax)_n = x_1^t (\sum_{j=1}^n x^t A_j x_j)_i + \cdots + x_n^t (\sum_{j=1}^n x^t A_j x_j)_n =$$

$$x_1^t (A_1 x_1 + A_2 x_2 + \cdots + A_n x_n) + \cdots + x_n^t (A_1 x_1 + A_2 x_2 + \cdots + A_n x_n) =$$

$$\begin{gathered} x_1^t A_{11} x_1 + x_1^t A_{12} x_2 + \cdots + x_1^t A_{1n} x_n + x_2^t A_{21} x_1 + x_2^t A_{22} x_2 + \cdots + \\ x_2^t A_{2n} x_n + \cdots + x_n^t A_{n1} x_1 + x_n^t A_{n2} x_2 + \cdots + x_n^t A_{nn} x_n = \\ \sum_{i=1}^n \sum_{j=1}^n A_{ij} x_i x_j = f(x) \end{gathered}$$

So when we take the derivative $\dfrac{\partial f}{\partial x_k}$, the terms where $x_i = x_k$ will be treated as constants then: $\dfrac{\partial f}{\partial x_1} x_1^t A_{11} x_1 = A_{11}$, $\dfrac{\partial f}{\partial x_2} x_1^t A_{12} x_2 = A_{12} x 2, \cdots, \dfrac{\partial f}{\partial x_n} x_n^t A_{nn} x_n = A_{nn}$

$$\nabla f(x) = \begin{bmatrix} \dfrac{\partial f}{\partial x_1} x_1^t A_{11} x_1 + x_1^t A_{12} x_2 + \cdots + x_n^t A_{nn} x_n \\ \vdots \\ \dfrac{\partial f}{\partial x_n} x_n^t A_{n1} x_1 + x_n^t A_{n2} x_2 + \cdots + x_n^t A_{nn} x_n \end{bmatrix}$$

Finally, we can express this as:

$$\sum_{i=1}^n x_i^t A_{ik} + \sum_{j=1}^n A_{jk} x_j = (A^t x)_k + (Ax)_k = ((A^t + A)x)_k$$

Therefore, $\nabla f(x) = (A^t + A)x$

(c) Since this is just a special case of the previous problem. We can say that $\nabla f(x) = (A^t + A)x$, where $A^t = A$, and both A^t and A are in $\mathbb{R}^{n\times n}$ then $\nabla f(x) = (A + A)x = 2Ax$

5. **Express the following statements in matrix language**You can assume that all matrices mentioned have appropriate dimensions. Here is an example: "Every column of C is a linear combination of the columns of B" can be expressed as "$C = BF$ for some matrix F".

   There can be several answers; one is good enough for us.

   (a) Suppose Z has n columns. For each i, row i of Z is a linear combination of rows $i, \cdots, n$ of Y.

   (b) W is obtained from V by permuting adjacent odd and even columns (i.e., 1 and 2, 3 and 4, $\cdots$).

   (c) Each column of P makes an acute angle with each column of Q

   (d) Each column of P makes an acute angle with the corresponding column of Q

   (e) The first k columns of A are orthogonal to the remaining columns of A.

   Solution 5. (a) $Z = AY$, where A is a matrix.

   (b) $W = VP$, where P is the permutation matrix

   (c) $P^tQ > 0$, then $\angle(P^tQ)$ is acute. (all entries are positive)

   (d) $p_i^t q_j > 0$, for all $i = j$

   (e) $A \in \mathbb{R}^{n\times m}$, $A^{\perp}_{n\times k} A_{m\times(n-k)} = 0$

$$A^{\perp}_{n\times k} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mk} \end{bmatrix}, A_{m\times(n-k)} = \begin{bmatrix} a_{1(k+1)} & a_{1(k+2)} & \cdots & a_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m(k+1} & a_{m(k+2)} & \cdots & a_{mn} \end{bmatrix}$$