CPSC 340 Assignment 0 Solutions

Sugun Machipeddy, Student No:65753337

2018-01-10

1 Linear Algebra Review

For these questions you may find it helpful to review these notes on linear algebra:
`http://www.cs.ubc.ca/~schmidtm/Documents/2009_Notes_LinearAlgebra.pdf`

1.1 Basic Operations

Rubric: {reasoning:7}

Use the definitions below,

$$\alpha = 2, \quad x = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \quad y = \begin{bmatrix} 3 \\ 4 \\ 5 \end{bmatrix}, \quad z = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}, \quad A = \begin{bmatrix} 3 & 2 & 2 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix},$$

and use x_i to denote element i of vector x. Evaluate the following expresions (you do not need to show your work).

1. $\sum_{i=1}^{n} x_i y_i$ (inner product).
2. $\sum_{i=1}^{n} x_i z_i$ (inner product between orthogonal vectors).
3. $\alpha(x + y)$ (vector addition and scalar multiplication).
4. $\|x\|$ (Euclidean norm of x).
5. x^T (vector tranpose).
6. A^T (matrix transpose).
7. Ax (matrix-vector multiplication).

1.1.1 solutions

1. $\sum_{i=1}^{n} x_i y_i = 0 * 3 + 1 * 4 + 2 * 5 = 14$
2. $\sum_{i=1}^{n} x_i z_i = 0 * 1 + 1 * 2 - 1 * 2 = 0$
3. $\alpha(x + y) = \begin{bmatrix} 6 \\ 10 \\ 14 \end{bmatrix}$

4. $\|x\| = \sqrt{0^2 + 1^2 + 2^2} = \sqrt{5}$
5. $x^T = \begin{bmatrix} 0 & 1 & 2 \end{bmatrix}$
6. $A^T = \begin{bmatrix} 3 & 1 & 1 \\ 2 & 3 & 1 \\ 2 & 1 & 3 \end{bmatrix}$
7. $Ax = \begin{bmatrix} 6 \\ 5 \\ 7 \end{bmatrix}$

1.2 Matrix Algebra Rules

Rubric: {reasoning:9}

Assume that $\{x, y, z\}$ are $n \times 1$ column vectors and $\{A, B, C\}$ are $n \times n$ real-valued matrices. State whether each of the below is true in general (you do not need to show your work).

1. $x^T y = \sum_{i=1}^n x_i y_i$.
2. $x^T x = \|x\|^2$.
3. $x^T(y + z) = z^T x + x^T y$.
4. $x^T(y^T z) = (x^T y)^T z$.
5. $AB = BA$.
6. $A(B + C) = AB + AC$.
7. $(AB)^T = A^T B^T$.
8. $x^T A y = y^T A^T x$.
9. $\det A \neq 0 \iff A$ is invertible.

1.2.1 solutions

1. true
2. true
3. false
4. false
5. false
6. true
7. false
8. true
9. true

1.3 Special Matrices

Rubric: {reasoning:3}

In one sentence, write down the defining properties of the following special types of matrices:

1. Symmetric matrix.
2. Identity matrix.
3. Orthogonal matrix.

1.3.1 solutions

1. Symmetric matrix : A symmetric matrix is a square matrix that is equal to its transpose ($A = A^T$).
2. Identity matrix : An identity matrix is a square matrix with ones in its diagonal and zeros else where.
3. Orthogonal matrix : An orthogonal matrix is a matrix whose transpose is equal to its inverse ($A^{-1} = A^T$).

2 Probability Review

For these questions you may find it helpful to review these notes on probability:
`http://www.cs.ubc.ca/~schmidtm/Courses/340-F15/notes_probability.pdf` And here are some slides giving visual representations of the ideas as well as some simple examples:
`http://www.cs.ubc.ca/~schmidtm/Courses/340-F17/probability.pdf`

2.1 Rules of probability

Rubric: {reasoning:6}

Answer the following questions. You do not need to show your work.

1. You flip 4 fair coins. What is the probability of observing exactly 3 heads?
2. You are offered the opportunity to play the following game: your opponent rolls 2 regular 6-sided dice. If the difference between the two rolls is at least 3, you win $12. Otherwise, you get nothing. What is a fair price for a ticket to play this game once? In other words, what is the expected value of playing the game?
3. Consider two events A and B such that $\Pr(A, B) = 0$. If $\Pr(A) = 0.4$ and $\Pr(A \cup B) = 0.95$, what is $\Pr(B)$? Note: $p(A, B)$ means "probability of A and B" while $p(A \cup B)$ means "probability of A or B". It may be helpful to draw a Venn diagram.

2.1.1 solutions

1. the probablitiy of observing exactly three heads is 0.25.
2. the probabality of winning is 1/3. Since the probability of winning is one in three chances, the fair price for a ticket should be $12/3 =$4.
3. $\Pr(A, B) = \Pr(A) + \Pr(B) - \Pr(A \cup B)$; $\Pr(B) = 0.55$

2.2 Bayes Rule and Conditional Probability

Rubric: {reasoning:10}

Answer the following questions. You do not need to show your work.

Suppose a drug test produces a positive result with probability 0.95 for drug users, $P(T = 1|D = 1) = 0.95$. It also produces a negative result with probability 0.99 for non-drug users, $P(T = 0|D = 0) = 0.99$. The probability that a random person uses the drug is 0.0001, so $P(D = 1) = 0.0001$.

1. What is the probability that a random person would test positive, $P(T = 1)$?
2. In the above, do most of these positive tests come from true positives or from false positives?
3. What is the probability that a random person who tests positive is a user, $P(D = 1|T = 1)$?
4. Are your answers from part 2 and part 3 consistent with each other?
5. What is one factor you could change to make this a more useful test?

2.2.1 solutions

1. the probability that a random person would test positive is the sum of probabilities of true positives and false positives.
   $P(T = 1) = P(T = 1|D = 1) * P(D = 1)) + P(T = 1|D = 0) * P(D = 0)$
   $P(T = 1) = 0.95 * 0.0001 + 0.01 * 0.9999 = 0.010094$
2. Most of the tests come from false positves. true positives account for $P(T = 1|D = 1) * P(D = 1)) = 0.000095$ and false positives account for $P(T = 1|D = 0) * P(D = 0) = 0.009999$. we can see that the result is dominated by false positives.
3. $P(D = 1|T = 1) = P(T = 1|D = 1) * P(D = 1))/P(T = 1) = 0.95 * 0.0001/0.010094 = 0.00941153$.
4. Yes, the answers from part 2 and part 3 are consistent with each other.
5. The probability of a random person using this drug is too small. this results in false positives dominating the result. Increasing the probability of a person using the drug would make the test more useful.

3 Calculus Review

3.1 One-variable derivatives

Rubric: {reasoning:8}

Answer the following questions. You do not need to show your work.

1. Find the minimum value of the function $f(x) = 3x^2 - 2x + 5$ for $x \in \mathbb{R}$.
2. Find the maximum value of the function $f(x) = x(1 - x)$ for x in the interval $[0, 1]$.
3. Find the minimum value of the function $f(x) = x(1 - x)$ for x in the interval $[0, 1]$.
4. Let $p(x) = \frac{1}{1+\exp(-x)}$ for $x \in \mathbb{R}$. Compute the derivative of the function $f(x) = -\log(p(x))$ and simplify it by using the function $p(x)$.

Remember that in this course we will $\log(x)$ to mean the "natural" logarithm of x, so that $\log(\exp(1)) = 1$. Also, obseve that $p(x) = 1 - p(-x)$ for the final part.

3.1.1 solutions

1. the minimum value of the funcion$f(x) is f(1/3) = 14/3$.
2. the maximum value of the function in the interval [0,1] is $f(1/2) = 1/4$.
3. the minimum value of the function in the interval [0,1] is $f(0) = f(1) = 0$.
4. the derivative of $f^{'}(x) = -\exp(-x)/(1 + \exp(-x))$.

3.2 Multi-variable derivatives

Rubric: {reasoning:5}

Compute the gradient $\nabla f(x)$ of each of the following functions. You do not need to show your work.

1. $f(x) = x_1^2 + \exp(x_2)$ where $x \in \mathbb{R}^2$.
2. $f(x) = \exp(x_1 + x_2 x_3)$ where $x \in \mathbb{R}^3$.
3. $f(x) = a^T x$ where $x \in \mathbb{R}^2$ and $a \in \mathbb{R}^2$.
4. $f(x) = x^\top A x$ where $A = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}$ and $x \in \mathbb{R}^2$.
5. $f(x) = \frac{1}{2}\|x\|^2$ where $x \in \mathbb{R}^d$.

Hint: it is helpful to write out the linear algebra expressions in terms of summations.

3.2.1 solutions

1. $f^{'}(x) = \begin{bmatrix} 2x_1 & exp(x_2) \end{bmatrix}$
2. $f^{'}(x) = \begin{bmatrix} \exp(x_1 + x_2 x_3) & x_3 * \exp(x_1 + x_2 x_3) & x_2 * \exp(x_1 + x_2 x_3) \end{bmatrix}$
3. $f^{'}(x) = a^T$
4. $f^{'}(x) = \begin{bmatrix} 4x_1 - 2x_2 & -2x_1 + 2x_2 \end{bmatrix} = 2x^T A$
5. $f^{'}(x) = x^T$

3.3 Derivatives of code

Rubric: {code:4}

Note: for info on installing and using Python, see `https://github.ugrad.cs.ubc.ca/CPSC340-2017W-T2/home/blob/master/python_info.md`.

Your repository contains a file named `grads.py` which defines several Python functions that take in an input variable x, which we assume to be a 1-d array (in math terms, a vector). It also includes (blank) functions that return the corresponding gradients. For each function, write code that computes the gradient of the function in Python. You can do this directly in `grads.py`; no need to make a fresh copy of the file. However, per the homework instructions, you should add a link to your README file so that the TA can access it easily.

Hint: it's probably easiest to first understand on paper what the code is doing, then compute the gradient, and then translate this gradient back into code.

Note: do not worry about the distinction between row vectors and column vectors here. For example, if the correct answer is a vector of length 5, we'll accept numpy arrays of shape `(5,)` (a 1-d array) or `(5,1)` (a column vector) or `(1,5)` (a row vector). In future assignments we will start to be more careful about this.

Warning: Python uses whitespace instead of curly braces to delimit blocks of code. Some people use tabs and other people use spaces. My text editor (Atom) inserts 4 spaces (rather than tabs) when I press the Tab key, so the file `grads.py` is indented in this manner. If your text editor inserts tabs, Python will complain and you might get mysterious errors... this is one of the most annoying aspects of Python, especially when starting out. So, please be aware of this issue! And if in doubt you can just manually indent with 4 spaces, or convert everything to tabs. For more information see `https://www.youtube.com/watch?v=SsoOG6ZeyUI`.

3.3.1 solutions

1. the derivatives can be found in grads.py file in code directory

4 Algorithms and Data Structures Review

4.1 Trees

Rubric: {reasoning:2}

Answer the following questions. You do not need to show your work.

1. What is the maximum number of *leaves* you could have in a binary tree of depth l?
2. What is the maximum number of *nodes* (including leaves) you could have in a binary tree of depth l?

Note: we'll use the convention that the leaves are not included in the depth, so for $l = 1$ the answers would be 2 and 3 respectively.

4.1.1 solutions

1. the maximum number of leaves in a binary tree of depth l is 2^l.
2. the maximum number of nodes in a binary tree of depth l is $2^{l+1} - 1$.

4.2 Common Runtimes

Rubric: {reasoning:4}

Answer the following questions using big-O notation You do not need to show your work.

1. What is the cost of running the mergesort algorithm to sort a list of n numbers?
2. What is the cost of finding the third-largest element of an unsorted list of n numbers?
3. What is the cost of finding the smallest element greater than 0 in a *sorted* list with n numbers.
4. What is the cost of computing the matrix-vector product Ax when A is $n \times d$ and x is $d \times 1$.

4.2.1 solutions

1. $O(n \log(n))$.
2. $O(n)$.
3. $O(n)$.
4. $O(nd1)$.

4.3 Running times of code

Rubric: {reasoning:4}

Your repository contains a file named `bigO.py`, which defines several functions that take an integer argument N. For each function, state the running time as a function of N, using big-O notation. Please include your answers in your report. Do not write your answers inside `bigO.py`.

4.3.1 solutions

1. $O(N)$.
2. $O(2N)$.
3. $O(1000N)$.
4. $O(N^2)$.