COURSEWORK

KEELE UNIVERSITY

SCHOOL OF COMPUTING AND MATHEMATICS

Mathematics for AI and Data Science

Author:
Sangeeth Dev Sarangi (CID: 21022161)

Date: November 17, 2021

1 Linear Algebra

In this section, we are going to solve following system of linear equations:

$$-3x + 2y - 6z = 6$$

$$5x + 7y - 5z = 6$$

$$x + 4y - 2z = 8$$

1.1 Solving the system of linear equations using Gaussian Elimination and Back Substitution.

- Consider the following system of 3 equations and 3 unknowns.

$$\begin{aligned} -3x + 2y - 6z &= 6 \\ 5x + 7y - 5z &= 6 \\ x + 4y - 2z &= 8 \end{aligned}$$

- If we multiply the second row with 3 and first row with 5 and add it on the second row we can eliminate x from the second row. We can use the notation $R_2 = 3R_2 + 5R_1$

$$\begin{aligned} -3x + 2y - 6z &= 6 \\ 5x + 7y - 5z &= 6 \\ x + 4y - 2z &= 8 \end{aligned} \xRightarrow{R_2=3R_2+5R_1} \begin{aligned} -3x + 2y - 6z &= 6 \\ 31y - 45z &= 48 \\ x + 4y - 2z &= 8 \end{aligned}$$

- Similarly if we multiply the third row with 3 and add first row on the third row we can eliminate x from the third row too.For this we can use the notation $R_3 = 3R_3 + R_1$

$$\begin{aligned} -3x + 2y - 6z &= 6 \\ 31y - 45z &= 48 \\ x + 4y - 2z &= 8 \end{aligned} \xRightarrow{R_3=3R_3+R_1} \begin{aligned} -3x + 2y - 6z &= 6 \\ 31y - 45z &= 48 \\ 14y - 12z &= 30 \end{aligned}$$

- Next step is to multiply the third row with 31 and second row with 14 and subtract it from the third row, so we eliminate y from the third equation.We can use the notation $R_3 = 31R_3 - 14R_2$

$$\begin{aligned} -3x + 2y - 6z &= 6 \\ 31y - 45z &= 48 \\ 14y - 12z &= 30 \end{aligned} \xRightarrow{R_3=31R_3-14R_2} \begin{aligned} -3x + 2y - 6z &= 6 \\ 31y - 45z &= 48 \\ 258z &= 258 \end{aligned}$$

- So far, we have produced an equivalent representation of the system of equations.

$$
\begin{aligned}
-3x + 2y - 6z &= 6 \\
5x + 7y - 5z &= 6 \\
x + 4y - 2z &= 8
\end{aligned}
\qquad
\begin{aligned}
-3x + 2y - 6z &= 6 \\
31y - 45z &= 48 \\
14y - 12z &= 30
\end{aligned}
\qquad
\begin{aligned}
-3x + 2y - 6z &= 6 \\
31y - 45z &= 48 \\
258z &= 258
\end{aligned}
$$

- It is often convenient to represent and operate on the augmented matrix form $[A|b]$, where ‘A’ is coefficient matrix of the LHS of the system of equations and ‘b’ is the RHS of system of equations.

$$
\left[\begin{array}{ccc|c}
-3 & 2 & -6 & 6 \\
5 & 7 & -5 & 6 \\
1 & 4 & -2 & 8
\end{array}\right]
\begin{array}{l}
R_2 = 3R_2 + 5R_1 \\
R_3 = 3R_3 + R_1
\end{array}
$$

$$
\rightsquigarrow
\left[\begin{array}{ccc|c}
-3 & 2 & -6 & 6 \\
0 & 31 & -45 & 48 \\
0 & 14 & -12 & 30
\end{array}\right]
R_3 = 31R_3 - 14R_2
$$

$$
\rightsquigarrow
\left[\begin{array}{ccc|c}
-3 & 2 & -6 & 6 \\
0 & 31 & -45 & 48 \\
0 & 0 & 258 & 258
\end{array}\right]
$$

- The solution to the system of linear equations after Gaussian Elimination, can be found by simply applying back-substitution.

$$
\begin{aligned}
-3x + 2y - 6z &= 6 \\
31y - 45z &= 48 \\
258z &= 258
\end{aligned}
\qquad
\begin{aligned}
& \\
& \\
z &= 1
\end{aligned}
\qquad
\begin{aligned}
& \\
y &= 3 \\
z &= 1
\end{aligned}
\qquad
\begin{aligned}
x &= -2 \\
y &= 3 \\
z &= 1
\end{aligned}
$$

- We can solve the equations in reverse order because the system after elimination is triangular.

$$
\begin{aligned}
258z &\Rightarrow 258 \\
z &\Rightarrow \frac{258}{258} \\
&\Rightarrow 1
\end{aligned}
\qquad
\begin{aligned}
31y - 45z &\Rightarrow 48 \\
y &\Rightarrow \frac{48 + 45(1)}{31} \\
&\Rightarrow 3
\end{aligned}
\qquad
\begin{aligned}
-3x + 2y - 6z &\Rightarrow 6 \\
x &\Rightarrow \frac{6 - 2(3) + 6(1)}{-3} \\
&\Rightarrow -2
\end{aligned}
$$

1.2 Solving the system of linear equations using python function 'numpy.linalg.solve'.

- We are considering the following system of 3 equations and 3 unknowns.

$$
\begin{aligned}
-3x + 2y - 6z &= 6 \\
5x + 7y - 5z &= 6 \\
x + 4y - 2z &= 8
\end{aligned}
$$

- Code snippet and execution results *(Platform used : Google Colab)*

Maths_coursework.ipynb

File Edit View Insert Runtime Tools Help Saving...

\+ Code + Text

```python
#Code Snippet
import numpy as np
a = np.array([[-3,2,-6],[5,7,-5],[1,4,-2]])
b = np.array([6,6,8])
x = np.linalg.solve(a, b)
print("OUTPUT :")
print("--------")
print('The array displays the value of x,y,z respectively - ',x)
print('Value of x = ',int(x[0]))
print('Value of y = ',int(x[1]))
print('Value of z = ',int(x[2]))
```

```
OUTPUT :
--------
The array displays the value of x,y,z respectively -  [-2.  3.  1.]
Value of x =  -2
Value of y =  3
Value of z =  1
```

- Varified that the results obtained using Gaussian Elimination with Back Substitution and python function - 'numpy.linalg.solve' are same for the given system of linear equations.

2 Optimization

2.1 Determining maximum and minimum value of a function

- Consider the function $f(x) = 3x^2 - 2x + 5$ in the interval [0,3]

- We will use the closed interval method to find the absolute maximum and minimum values of a continuous function f on the closed interval *[a,b]*.

- For that, first we need to find the values of f at the critical numbers of f in *(a,b)* ,then we will find the values of f at the endpoints of the interval.The largest and smallest value from this set are the absolute maximum and absolute minimum respectively.

$$f(x) = 3x^2 - 2x + 5$$

$$f'(x) = 6x - 2$$

- Since, $f'(x)$ exists for all x, the only critical point of f occurs when $f'(x) = 0$

$$\begin{aligned} f'(x) &\Rightarrow 0 \\ 6x - 2 &\Rightarrow 0 \\ x &\Rightarrow \frac{2}{6} \\ &\Rightarrow \frac{1}{3} \end{aligned}$$

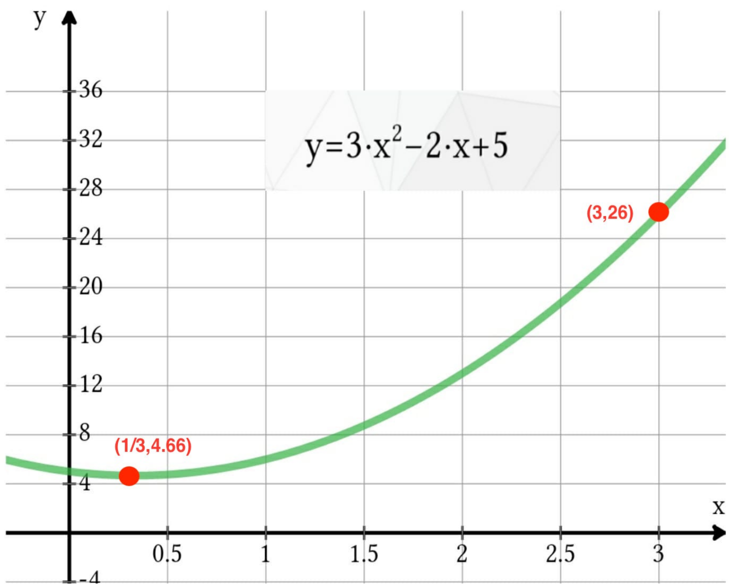
- Value of f at critical points is :$f(\frac{1}{3}) = \frac{14}{3} \Rightarrow 4.66$

- The value of f at the endpoints of the interval are :

$$f(0) \Rightarrow 5 \qquad\qquad f(3) \Rightarrow 26$$

- Comparing these 3 values, we can see that the absolute maximum value is $f(3) = 26$ and the absolute minimum value is $f(\frac{1}{3}) = 4.66$

- In this problem, we could see, the absolute maximum occurs at an endpoint, whereas the absolute minimum occurs at a critical number. The graph of f is sketched below for reference :



2.2 Concept of the gradient descent method with a help of a suitable example

- Gradient Descent is an optimization algorithm for finding a local minimum of a differentiable function. Gradient descent is simply used in machine learning to find the values of a function's parameters (coefficients) that minimize a cost function as far as possible.

- Imagine a blindfolded man is in the top of a mountain and he is attempting to reach the bottom field.In order to achieve this goal, he needs to take small steps around and move towards the direction of the higher incline. He should do this iteratively, moving one step at a time until finally reach the bottom of the mountain.Gradient Descent accomplishes exactly that. Its purpose is to get to the mountain's lowest point. The mountain is the data plotted in space, the learning rate is the size of the step you take, and calculating the gradient of a set of parameter values, which is done iteratively, is feeling the incline around you and deciding which is higher. The preferred direction is one in which the cost function decreases (the opposite direction of the gradient). The lowest point on the mountain is the value -or weights- at which the function's cost is at its cheapest (the parameters where our model presents more accuracy).

- The equation below describes what gradient descent does: x is the next position of our climber, while a represents his current position. The minus sign refers to the minimization part of gradient descent. The α in the middle is a waiting factor and the gradient term $f'(x)$ is simply the direction of the steepest descent.

$$x = x - \alpha f'(x)$$

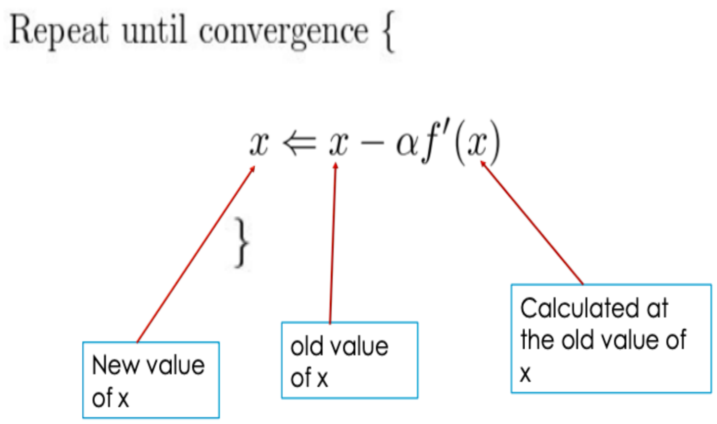
- So, basically, this formula informs us the position that we need to go next, which is in the direction of the steepest descent.

2.3 Advantage of the Gradient descent method as compared to the First derivative test for finding the maximum and minimum of a function

- Finding optimal value using derivative is difficult for complex function.That is,when the target function is extremely difficult, calculating the root of the derivative of a function would be incredibly hard. More crucially,we must teach the computer how to determine the root of a function's derivative.Knowing the explicit statement of the root, such as the derivative of x^2, would make our lives much easier.However, if the target function appears illogical (has no closed form), we should devise a specific technique to teach the computer that this is the best way to obtain the (local) optimal answer. The algorithm that the computer understands in this case is gradient descent.It is very popular approach for optimizing objective functions (irrespective of their functional form).

- Solving an equation for zero in 4-space and higher dimensions or for large systems of equations can be quite challenging. Approximation by employing gradient descent may be much faster than trying to solve the problem by hand or with a computer.

- In gradient descent, one starts at an initial point $x = x_0$ and successively updates x using the steepest descent direction. Here, $\alpha > 0$ regulates the step size,and also referred to as the learning rate.We should also consider the fact that it is an iterative algorithm



- Performing this update again and again to construct the sequence $x_0, x_1, x_2, ..x_t$ and it will eventually converge to the optimal value.

2.4 Gradient descent method to find the minimum of the function

- Consider the function $f(x) = x^2$. Assume that the learning rate $\alpha = 0.1$ and the initial value is $x = 2$.Using this information, we are going to show the first 3 iterations of the application of the gradient descent method to find the minimum of the function.

$$Function : f(x) = x^2$$

$$Starting_Point : x_0 = 2$$

$$Learning_Rate : \alpha = 0.1$$

- Iteration 1 :

$$f'(x) \Rightarrow 2x$$
$$f'(2) \Rightarrow 2 * 2 = 4$$
$$x_1 \Leftarrow x_0 - 0.1 * 4 = 2 - 0.4 = 1.6 \qquad \sim x = x - \alpha f'(x)$$

- Iteration 2 :

$$x_2 \Leftarrow x_1 - 0.1 * f'(1.6) \qquad \sim x = x - \alpha f'(x)$$

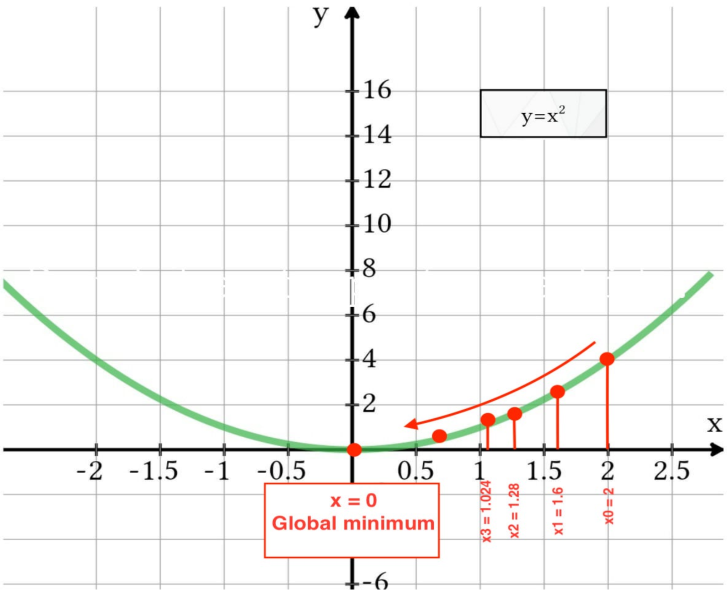
$$x_2 \Leftarrow 1.6 - 0.1 * (2 * 1.6) = 1.6 - 0.32 = 1.28$$

- Iteration 3 :

$$x_3 \Leftarrow x_2 - 0.1 * f'(1.28) \qquad\qquad \sim x = x - \alpha f'(x)$$

$$x_3 \Leftarrow 1.28 - 0.1 * (2 * 1.28) = 1.28 - 0.256 = 1.024$$

- Performing this update again and again to construct the sequence $x_0, x_1, x_2, ..x_t$ and it will eventually converge to the optimal value of $x_t = 0$ for large values of t.The graph of f is sketched below for reference :



3 Conditional probability

3.1 Concept of the Bayes theorem of conditional probability

- Bayes theorem is a way to figure out conditional probability. Conditional probability is the probability of an event happening, given that it has some relationship to one or more other events.

- Bayes Theorem (also known as Bayes rule) is a deceptively simple formula used to calculate conditional probability.The formal definition for the rule is:

$$P(A_i|B) = \frac{P(A_i)P(B|A_i)}{P(B)}$$
$$= \frac{P(A_i)P(B|A_i)}{P(A_1)P(B|A_1)+...+P(A_n)P(B|A_n)}$$

- Where A and B are two events and $P(B) \neq 0$

  - $A1, A2, ..., An$ are the partition of the sample space. Also $P(A_i) > 0$ for all i
  - $P(A|B)$ is the conditional probability of event A occurring given that B is true
  - $P(B|A)$ is the conditional probability of event B occurring given that A is true.
  - $P(A)$ and $P(B)$ are the probabilities of A and B occurring independently of one another.

- Bayes Theorem Example - Liver Disease Detection

  - In this example, we are going to find out a patient's probability of having liver disease if they are an alcoholic. The test (sort of like a litmus test) for liver disease is 'being an alcoholic'. According to previous data, 20 % of patients admitted to the hospital have liver disease.6 % of the hospital's patients are alcoholics.We might also know the fact that among those patients diagnosed with liver disease, 8 % are alcoholics.
  - Let A be the event says 'Patient has liver disease'.
    $P(A) = 0.20$
  - Let B mean the litmus test that 'Patient is being an alcoholic'.
    $P(B) = 0.06$
  - The probability that a patient is alcoholic, given that they have liver disease, is 8%
    $P(B|A) = 0.08$
  - By applying Bayes theorem :

$$P(A|B) = \frac{P(A)*P(B|A)}{P(B)}$$

$P(A|B) = \frac{0.20*0.08}{0.06} = 0.2666$

  - In other words, if the patient is an alcoholic, their chances of having liver disease is 0.2666 (ie; 26.66%)

3.2 Computing the following conditional probabilities

- Table 1 shows the feature table, i.e., features of fruits selected for feature measurement. Suppose that a fruit is selected randomly.Using the information given in Table 1, we are going to compute the following probabilities.

Type	Long	Not Long	Sweet	Not Sweet	Yellow	Not Yellow	Total
Banana	400	100	350	150	450	50	500
Orange	0	300	150	150	300	0	300
Other	100	100	150	50	50	150	200
Total	500	500	650	350	800	200	1000

Table 1: Features of fruits selected for feature measurement

1. P(Banana) $= \frac{500}{1000} = 0.5$
2. P(Orange) $= \frac{300}{1000} = 0.3$
3. P(Other) $= \frac{200}{1000} = 0.2$

3.3 Computing the following conditional probabilities

- Using the information given in the above table, i.e.,'Table 1:Features of fruits selected for feature measurement'. We are going to compute the following conditional probabilities.

1. P(Long | Banana) $= \frac{400}{500} = 0.8$
2. P(Sweet | Banana) $= \frac{350}{500} = 0.7$
3. P(Yellow | Banana) $= \frac{450}{500} = 0.9$
4. P(Long | Orange) $= \frac{0}{300} = 0$
5. P(Sweet | Orange) $= \frac{150}{300} = 0.5$
6. P(Yellow |Orange) $= \frac{300}{300} = 1$