Lecture A1. Math Review

Sim, Min Kyu, Ph.D., mksim@seoultech.ac.kr

SEOULTECH 서울과학기술대학교 데이터사이언스학과

I. Differentiation and integration

Differentiation

Definition 1 (differentiation)

Differentiation is the action of computing a derivative.

Definition 2 (derivative)

The **derivative** of a function $y = f(x)$ of a variable x is a measure of the rate at which the value y of the function changes with respect to (wrt., hereafter) the change of the variable x. It is notated as $f'(x)$ and called **derivative** of f wrt. x.

Remark 1

If x and y are real numbers, and if the graph of f is plotted against x, the derivative is the **slope of this graph** at each point.

Definition 3 (differentiable)

If $\lim_{h\to 0} \frac{f(x+h/2)-f(x-h/2)}{h}$ exists for a function f at x, we say the function f **is differentiable at** x. That is, $f'(x) = \lim_{h\to 0} \frac{f(x+h/2)-f(x-h/2)}{h}$. If f is differentiable for all x, then we say f **is differentiable (everywhere)**.

Remark 2

The followings are popular derivatives.

- $f(x) = x^p \Rightarrow f'(x) = px^{p-1}$ (polyomial)
- $f(x) = e^x \Rightarrow f'(x) = e^x$ (exponential)
- $f(x) = log(x) \Rightarrow f'(x) = 1/x$ (log function; not differentiable at $x = 0$)

Theorem 1

Differentiation is linear. That is, $h(x) = f(x) + g(x)$ *implies* $h'(x) = f'(x) + g'(x)$.

Theorem 2 (differentiation of product)

If $h(x) = f(x)g(x)$, *then* $h'(x) = f'(x)g(x) + f(x)g'(x)$.

Exercise 1

Suppose $f(x) = xe^x$, *find* $f'(x)$.

Theorem 3 (differentiation of fraction)

If $h(x) = \frac{f(x)}{g(x)}$, *then* $h'(x) = \frac{f'(x)g(x) - f(x)g'(x)}{(g(x))^2}$.

Theorem 4 (composite function)

If $h(x) = f(g(x))$*, then* $h'(x) = f'(g(x)) \cdot g'(x)$.

Exercise 2

Suppose $f(x) = e^{2x}$*, find* $f'(x)$.

Differentiation

- Oftentimes, finding analytic derivative is hard, but finding numerical derivative is often possible.

Definition 4

For a function f and a small constant h,

- $f'(x) \approx \frac{f(x+h)-f(x)}{h}$ *(forward difference formula)*
- $f'(x) \approx \frac{f(x)-f(x-h)}{h}$ *(backward difference formula)*
- $f'(x) \approx \frac{f(x+h)-f(x-h)}{2h}$ *(centered difference formula)*

Integration

Definition 5 (integration)

Integration is the computation of an integral, which is a reverse operation of differentiation up to an additive constant.

Definition 6 (integral or antiderivative)

Let's say a function f is a derivative of g, or $g'(x) = f(x)$, then we say g is an **integral** or, **antiderivative** of f, written as $g(x) = \int f(x)dx + C$, where C is a integration constant.

Remark 3

The followings are popular integrals.

- For $p \neq 1$, $f(x) = x^p \Rightarrow \int f(x)dx = \frac{1}{p+1}x^{p+1} + C$ (polyomial)
- $f(x) = \frac{1}{x} \Rightarrow \int f(x)dx = log(x) + C$ (fraction)
- $f(x) = e^x \Rightarrow \int f(x)dx = e^x + C$ (exponential)
- $f(x) = \frac{g'(x)}{g(x)} \Rightarrow \int f(x)dx = log(g(x)) + C$ (See Theorem 4 above)

Exercise 3

Derive $\int f'(x)g(x)\ dx = f(x)g(x) - \int f(x)g(x)'\ dx$. *(Hint: Use Theorem 2 above.)*

Exercise 4

Find $\int xe^x \ dx$, *and evaluate* $\int_0^1 xe^x \ dx$. *(Hint: Use Exercise 3 above.)*

II. Numerical methods for finding a root

About - solving an equation

- For the rest of this section, we consider a nonlinear and differentiable (thus, continuous) function $f : \mathbb{R} \to \mathbb{R}$, we aim to find a point $x^* \in \mathbb{R}$ such that $f(x^*) = 0$. We call such x^* as a *solution* or a *root*.

1. Bisection Method

- The *bisection* method aims to find a very short interval $[a, b]$ in which f changes a sign.
- Why? Changing a sign from a to b means the function crosses the $\{y = 0\}$-axis (in other words x-axis) at least once. It means that x^* such that $f(x^*) = 0$ is within this interval. Since $[a, b]$ is a very short interval, We may simply say $x^* = \frac{a+b}{2}$.

Definition 7 (sign function)

$sgn(\cdot)$ is called a *sign function* that returns 1 if the input is positive, -1 if negative, and 0 if zero.

- Let tol be the maximum allowable length of the **short interval** and an initial interval $[a, b]$ be such that $sgn(f(a)) \neq sgn(f(b))$.
- The **bisection algorithm** is the following.

while $((b-a) > tol)$ do
    $m = \frac{a+b}{2}$
    if $sgn(f(a)) = sgn(f(m))$ then
        $a = m$
    else
        $b = m$
    end
end

- At each **iteration**, the interval length is halved. As soon as the interval length becomes smaller than tol, then the algorithm stops.

2. Newton Method

- The bisection technique makes no use of the function values other than their signs, resulting in slow but sure convergence.
- More rapid convergence can be achieved by using the function values to obtain a more accurate approximation to the solution, at each iteration.
- Newton method is a method that use both the function value and derivative value.
- Newton method approximates the function f near x_k by the tangent line at $f(x_k)$.

1: x_0 = initial guess
2: for $k = 0, 1, 2, \ldots$
3: $\quad x_{k+1} = x_k - f(x_k)/f'(x_k)$
4: $\quad$ break if $|x_{k+1} - x_k| < tol$
5: end

3. Fixed point theorem

Definition 8 (Fixed point)

For a function $f(\cdot)$, x^* is called a fixed point if $f(x^*) = x^*$ holds.

Remark 4

- For example, $x^* = 2$ is a fixed point for $f(x) = x^2 - 3x + 4$.
- Not all functions have fixed points. For example, $f(x) = x + 1$.
- In graphical terms, a fixed point x means the point $(x, f(x))$ is on the line $y = x$.
- In other words the graph of f has a point in common with that line.

Theorem 1 (contraction mapping theorem)

Let x_0 to be an arbitrary point, and let $x_{k+1} = f(x_k)$ for $k \geq 0$. Under certain condition of f, the sequence of $\{x_n\}$ converges to x^* such that $f(x^*) = x^*$.

$x_0 =$ initial guess
for $k = 0, 1, 2, \ldots$
    $x_{k+1} = x_k - f(x_k)/f'(x_k)$
    break if $|x_{k+1} - x_k| < tol$
end

- Consider $f(x) = 1 + 1/x$.
- Its solution to $f(x^*) = x^*$ can be solved by $x = 1 + 1/x$, or $x^2 - x - 1 = 0$.
- In other words, $x^* = \frac{1 \pm \sqrt{5}}{2} \approx 1.618 \; or \; -0.618$

```
f <- function(x) {
  return(1+1/x)
}
tol <- 10^(-5)
x_now <- 0.1
```

```
repeat{
  x_next <- f(x_now)
  if (abs(x_next-x_now) < tol) {
    break
  }
  x_now <- x_next
  print(x_next)
}
```

```
## [1] 11
## [1] 1.090909
## [1] 1.916667
## [1] 1.521739
## [1] 1.657143
## [1] 1.603448
## [1] 1.623656
## [1] 1.615894
## [1] 1.618852
## [1] 1.617722
## [1] 1.618153
## [1] 1.617988
## [1] 1.618051
## [1] 1.618027
```

Exercise 5

Write a python code that does the exactly same thing as the above code block.

Summary

- The above mentioned root-finding numerical methods share a few common properties.
  1. It is characterized as a *iterative process* (such as $x_0 \to x_1 \to x_2 \to \cdots$).
  2. In each *iteration*, the current candidate for the solution *gets closer* to the true value.
  3. It converges. That is, it is theoretically reach the *exact value* up to tolerance.

- Many iterative numerical methods share the properties above.
- The famous back propagation in deep neural network is also motivated by Newton method.
- Major algorithms for dynamic programming are called *policy iteration* and *value iteration* that also share the properties above.

III. Matrix algebra

Matrix multiplication

Exercise 6

Solve the followings.

$$\begin{pmatrix} .6 & .4 \end{pmatrix} \begin{pmatrix} .7 & .3 \\ .5 & .5 \end{pmatrix} =$$

Exercise 7

What is P^2?

$$P = \begin{pmatrix} .7 & .3 \\ .5 & .5 \end{pmatrix}$$

Solution to system of linear equations

Exercise 8

Solve the followings.

$$\begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{pmatrix} \begin{pmatrix} .7 & .3 \\ .5 & .5 \end{pmatrix} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 \end{pmatrix}$$

$$\mathbf{v}_1 + \mathbf{v}_2 = 1$$

Exercise 9

Solve the following system of equations.

$$
\begin{aligned}
x &= y \\
y &= 0.5z \\
z &= 0.6 - 0.4x \\
x + y + z &= 1
\end{aligned}
$$

Exercise 10

Solve the following system of equations.

$$\begin{pmatrix} \mathbf{v}_0 & \mathbf{v}_1 & \mathbf{v}_2 \end{pmatrix} \begin{pmatrix} -2 & 2 & \\ 3 & -5 & 2 \\ & 3 & -3 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \end{pmatrix}$$

$$\mathbf{v}_0 + \mathbf{v}_1 + \mathbf{v}_2 = 1$$

Exercise 11

Solve the following system of equations.

$$\begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4 \end{pmatrix} \begin{pmatrix} .7 & .3 & & \\ .5 & .5 & & \\ & & .6 & .4 \\ & & .3 & .7 \end{pmatrix} = \begin{pmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 & \mathbf{v}_4 \end{pmatrix}$$

$$\begin{aligned} \mathbf{v}_1 + \mathbf{v}_2 + \mathbf{v}_3 + \mathbf{v}_4 &= 1 \\ \mathbf{v}_1 + \mathbf{v}_2 &= a \end{aligned}$$

Exercise 12

Solve following and express $\mathbf{v}_i$ *for* $i = 0, 1, 2, \ldots$

$$
\begin{aligned}
\mathbf{v}_0 + \mathbf{v}_1 + \mathbf{v}_2 + \ldots &= 1 \\
0.02\mathbf{v}_0 + 0.02\mathbf{v}_1 + 0.02\mathbf{v}_2 + \ldots &= \mathbf{v}_0 \\
0.98\mathbf{v}_0 &= \mathbf{v}_1 \\
0.98\mathbf{v}_1 &= \mathbf{v}_2 \\
0.98\mathbf{v}_2 &= \mathbf{v}_3 \\
\ldots &= \ldots
\end{aligned}
$$

IV. Series and others

Exercise 13 (Infinite geometric series)

Simplify the following. When $|r| < 1$, $S = a + ar + ar^2 + ar^3 + ...$

Exercise 14 (Finite geometric series)

Simplify the following. When $r \neq 1$, $S = a + ar + ar^2 + ar^3 + ... + ar^{n-1}$

Exercise 15 (Power series)

Simplify the following. When $|r| < 1$, $S = r + 2r^2 + 3r^3 + 4r^4 + \ldots$

Formulation of time varying function

Exercise 16

During the first hour $(0 \leq t \leq 1)$, $\lambda(t)$ *increases linearly from* 0 *to* 60*. After the first hour,*$\lambda(t)$ *is constant at* 60*. Draw plot for* $\lambda(t)$ *and express the function in math form.*

```
"Man can learn nothing unless he proceeds from the known to the unknown. - Claude Bernard"
```