Lecture 7

Math Foundations Team

BITS Pilani
Pilani | Dubai | Goa | Hyderabad

Introduction

- In last lecture, we discussed about differentiation of univariate functions, partial differentiation, gradients and gradients of vector valued functions, and gradient of Matrices
- In this lecture we will study the Taylor Series expansion of a function and Taylor Polynomial expansion of a function
- We will also study the Hessian matrix, and how to find local maxima minima and Absolute maxima and minima

Applications of Taylor Series in ML

1. One of the most common applications of the Taylor Series in ML is in optimization problems. In optimizing a cost function, Taylor series can be used to approximate this function, making it easier to calculate gradients and perform optimizations.

2. The Taylor series can approximate complex functions in simpler polynomial forms, which is particularly useful in regression.

3. The Taylor series can also be used to understand and interpret the behaviour of ML models. By expanding the model's function around a point, we can gain insights into how changes in input affect the output, which is crucial for tasks like feature importance analysis and debugging models.

Taylor polynomial

The Taylor polynomial is a representation of a function f as an finite sum of terms. These terms are determined using derivatives of f evaluated at x_0.

Definition: The Taylor polynomial of degree n of $f : \mathbb{R} \to \mathbb{R}$ at x_0 is defined as

$$T_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k \tag{4}$$

where $f^{(k)}(x_0)$ is the kth derivative of f at x_0 which we assume exists.

Taylor polynomial

The Taylor polynomial is a representation of a function f as an finite sum of terms. These terms are determined using derivatives of f evaluated at x_0.

Definition: The Taylor polynomial of degree n of $f : \mathbb{R} \to \mathbb{R}$ at x_0 is defined as

$$T_n(x) = \sum_{k=0}^{n} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k \tag{4}$$

where $f^{(k)}(x_0)$ is the kth derivative of f at x_0 which we assume exists.

Taylor series

Definition: The Taylor series of smooth (continuously differentiable infinite many times) function $f : \mathbb{R} \to \mathbb{R}$ at x_0 is defined as

$$T_\infty(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k \tag{5}$$

For $x_0 = 0$, we obtain the Maclaurin series as a a special instance of the Taylor series.

Remark: In general, a Taylor polynomial of degree n is an approximation of a function, which does not need to be a polynomial. The Taylor polynomial is similar to f in a neighborhood around x_0. However, a Taylor polynomial of degree n is an exact representation of a polynomial f of degree $k \leq n$ since all derivatives $f^{(i)} = 0$, for $i > k$.

Taylor Polynomial example

Consider the polynomial $f(x) = x^4$. Find the Taylor polynomial T_6 evaluated at $x_0 = 1$.
We compute $f^{(k)}(1)$ for $k = 0, 1, 2..., 6$
$f(1) = 1$, $f'(1) = 4$, $f''(1) = 12$, $f^{(3)}(1) = 24$, $f^{(4)}(1) = 24$, $f^{(5)}(1) = 0$, $f^{(6)}(1) = 0$. The desired Taylor polynomial is

$$\begin{aligned} T_6(x) &= \sum_{k=0}^{6} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k \\ &= 1 + 4(x-1) + 12(x-1)^2 + 24(x-1)^3 + 24(x-1)^4 \\ &= x^4 = f(x) \end{aligned} \tag{6}$$

we obtain an exact representation of the original function.

Taylor Series example

Consider the smooth function $f(x) = sin(x) + cos(x)$. We compute Taylor series expansion of f at $x_0 = 0$, which is the Maclaurin series expansion of f. We obtain the following derivatives:

$f(0) = sin(0) + cos(0) = 1$

$f^{'}(0) = cos(0) - sin(0) = 1$

$f^{''}(0) = -sin(0) - cos(0) = -1$

$f^{(3)}(0) = -cos(0) + sin(0) = -1$

$f^{(4)}(0) = sin(0) + cos(0) = f(0) = 1$

The coefficients in our Taylor series are only ± 1 (since $sin(0) = 0$), each of which occurs twice before switching to the other one. Furthermore, $f^{(k+4)}(0) = f^{k}(0)$

Taylor Series example

Therefore, the full Taylor series expansion of f at $x_0 = 0$ is given by

$$\begin{aligned}
T_\infty(x) &= \sum_{k=0}^{\infty} \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k \\
&= 1 + x - \frac{1}{2!}x^2 - \frac{1}{3!}x^3 + \frac{1}{4!}x^4 + \frac{1}{5!}x^5 - \dots \\
&= 1 - \frac{1}{2!}x^2 + \frac{1}{4!}x^4 \mp \dots x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 \mp \dots \qquad (7) \\
&= \sum_{k=0}^{\infty}(-1)^k \frac{1}{(2k)!}x^{2k} + \sum_{k=0}^{\infty}(-1)^k \frac{1}{(2k+1)!}x^{2k+1} \\
&= cos(x) + sin(x)
\end{aligned}$$

Where does the Taylor series come from?

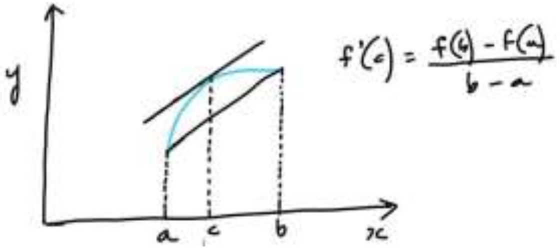
- Our development of the Taylor series will mirror the argument given in the document "Proof of Taylor's theorem" which will also be uploaded as class material.

  Theorem: Suppose $f:(a,b)\to R$ is a function on (a,b), where a,b in R with $a<b$. Assume that f is n-times differentiable in the open interval (a,b) and $f, f', f'', \ldots f^{n-1}$ all extend continuously to the closed interval $[a,b]$, such that the extended functions are still called $f, f'' \ldots f^{n-1}$. Then there exists $c\in(a,b)$ such that

$$f(b)=\sum_{k=0}^{k=n-1}\frac{f^{k}(a)}{k!}(b-a)^{k}+\frac{f^{n}(c)}{n!}(b-a)^{n} \qquad (1)$$

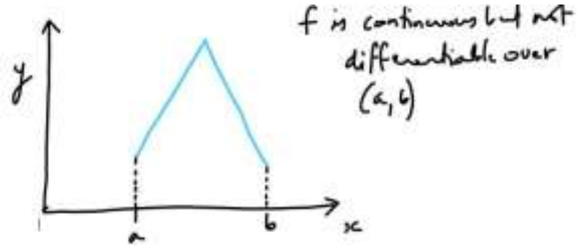
Mean-value theorem

- For $n = 1$, the statement of Taylor's theorem boils down to the mean-value theorem which is that if a function f is continuous on $[a, b]$ and differentiable on the interval (a, b), then there exists a value $c \in (a, b)$ such that $f'(c) = \frac{f(b) - f(a)}{b - a}$ as in the following figure:



Mean-value theorem

- Note that the requirement that f be a differentiable function in the mean-value theorem is needed as the theorem is not valid for functions f that are not differentiable as in the example below:



Mean-value theorem

- The proof of the mean-value theorem comes from Rolle's theorem whose statement follows. We shall show that the development of Taylor's series involves the repeated application of Rolle's theorem below:
  Theorem: If f is a continuous function on $[a, b]$ **and differentiable on** (a, b) **with** $f(a) = f(b) = 0$**, then there exists some** c **in** (a, b) **such that** $f'(c) = 0$**.**

Development of Taylor's series

- Let $F(x)$ be a function over the region $(a, b) \in R$ such that $F(a) = F'(a) = F^2(a) = \ldots = F^{n-1}(a) = 0$, and $F(b) = 0$. Then there exists a $c \in (a, b)$ such that $F^n(c) = 0$. Let us call this Proposition P.
- Proposition P follows from an n-fold application of Rolle's theorem as follows: since $F(a) = F(b) = 0$, an application of Rolle's theorem tells us that there is a $c_1 \in (a, b)$ such that $F'(c_1) = 0$.
- Now since $F'(a) = F'(c_1) = 0$, there exists $c_2 \in (a, c_1)$ such that $F''(c_2) = 0$.
- Continuing this argument we get $a < c_n < c_{n-1} < \ldots < c_1 < b$ such that $F^k(c_k) = 0$ for $k = 1, 2, \ldots n$.

Development of Taylor's series

- Thus we have $F^n(c) = 0$ for $c = c_n \in (a, b)$.
- To construct a polynomial that approximates a function f we use the ideas of the previous slide.
- Let the polynomial used to approximate the function f be of the form $P(x) = \sum_{k=0}^{k=n} a_k (x-a)^k$. We will now find the coefficients $a_0, a_1, a_2, \ldots a_n$ such that $F(x) = f(x) - P(x)$ satisfies $F(a) = F(b)$ and $F(a) = F'(a) = F^2(a) = \ldots = F^{n-1}(a) = 0$, and $F(b) = 0$. Then we have $F^n(c) = 0$ for $c \in (a, b)$.
- We can see that $F^k(a) = f^k(a) - k! a_k$, $k = 0, 1, 2, \ldots n-1$, and to satisfy the requirements of the function F defined on the previous slide we need to set $F(a) = F'(a) = F^{n-1}(a) = 0$, and $F(b) = 0$. This gives $a_k = \frac{f^k(a)}{k!}$, $k = 0, 1, 2, \ldots n-1$.

Development of Taylor's series

- We finally need to determine a_n. To do so, we note that $F(b) = f(b) - \sum_{k=0}^{k=n} a_k(b-a)^k = 0$. This becomes $F(b) = f(b) - \sum_{k=0}^{k=n-1} \frac{f^k(a)}{k!}(b-a)^k - a_n(b-a)^n = 0$ once we substitute for $a_k = \frac{f^k(a)}{k!}$, for $k = 1, 2, \ldots n-1$.
- We can then solve for a_n to get $a_n = \frac{1}{(b-a)^n}(f(b) - \sum_{k=0}^{k=n-1} \frac{f^k(a)}{k!}(b-a)^k)$.
- Our function F satisfies the requirements of proposition P, i.e ($F(a) = F(b)$ and $F'(a) = F''(a) = \ldots F^{n-1}(a) = 0$). Therefore there exists a point c, according to proposition P such that $F^n(c) = 0$.

Development of Taylor's series

- Since $F^n(c) = f^n(c) - P^n(c)$, we have
  $F^n(c) = f^n(c) - n!a_n =$
  $f^n(c) - \frac{n!}{(b-a)^n}(f(b) - \sum_{k=0}^{k=n-1} \frac{f^k(a)}{k!}(b-a)^k) = 0,$
- Rearranging the above we have the final Taylor's series expansion with a remainder term:
  $f(b) = \sum_{k=0}^{k=n-1} \frac{f^k(a)}{k!}(b-a)^k + \frac{f^n(c)}{n!}(b-a)^n.$
- The last term in the expression above is the remainder term which should be used when the series is truncated at a certain number of terms.

Taylor's series in two variables

- How do we develop Taylor's series in two variables?
- Let $f(x, y)$ be a function in two variables with continuous partial derivatives in an open region R containing the point $P(a, b)$ where the partial derivates f_x and f_y are both zero. Note that f_x and f_y are zero because the gradient vanishes at critical points, and (a, b) is a critical point.
- Let h and k be increments small enough to put the point $S(a+h, b+k)$ in the region R. We parameterise the line segment PS as $x = a + th, y = b + tk, t \in [0, 1]$.
- Now let $F(t) = f(a+th, b+tk)$. F is now a function of only one variable. We can compute $F'(t) = f_x \frac{dx}{dt} + f_y \frac{dy}{dt} = hf_x + kf_y$.

Taylor's series in two variables

- Now f_x and f_y are differentiable functions, F' is a differentiable function of t and we can write $F''(t) = \frac{\partial F'}{\partial x}\frac{dx}{dt} + \frac{\partial F'}{\partial y}\frac{dy}{dt}$.
- Since $x = a + th$ and $y = b + tk$, and $F' = hf_x + kf_y$, we can write
$F''(t) = \frac{\partial(hf_x+kf_y)}{\partial x}h + \frac{\partial(hf_x+kf_y)}{\partial y}k = h^2f_{xx} + 2hkf_{xy} + k^2f_{yy}$.
- Since F and F' are continuous on $[0, 1]$ and F' is differentiable on $(0, 1)$ we can apply Taylor's theorem and obtain $F(1) = F(0) + F'(0)(1 - 0) + \frac{1}{2}F''(c)$ for some c between 0 and 1.
- Rewriting this in terms of x, y, we have
$f(a + h, b + k) = f(a, b) + hf_x(a, b) + kf_y(a, b) + \frac{1}{2}(h^2f_{xx} + 2hkf_{xy} + k^2f_{yy})|_{a+ch,b+ck}$

Taylor's series in two variables

- Since $f_x(a,b) = f_y(a,b) = 0$, the expression for $f(a+h, b+k)$ simplifies to $f(a+h, b+k) - f(a,b) = \frac{1}{2}(h^2 f_{xx} + 2hkf_{xy} + k^2 f_{yy})|_{a+ch,b+ck}$
- The presence of an extremum at $f(a,b)$ is dependent on the sign of $f(a+h, b+k) - f(a,b)$ for arbitrary h and k.
- This is the same as the sign of $Q(c) = (h^2 f_{xx} + 2hkf_{xy} + k^2 f_{yy})|_{a+ch,b+ck}$.
- We shall now study the sign of $Q(c)$.

Hessian matrix

- If $Q(0) \neq 0$, the sign of $Q(c)$ for small c will be the same as the sign of $Q(0)$ for sufficiently small values of h and k.
- We can predict the sign of $Q(0) = (h^2 f_{xx}(a,b) + 2hkf_{xy}(a,b) + k^2 f_{yy}(a,b))$ from the signs of f_{xx} and $f_{xx}f_{yy} - f_{xy}^2$ evaluated at (a,b).
- Multiply both sides of the equation for $Q(0)$ by f_{xx}, and rearrange the right hand side to get $f_{xx}Q(0) = (hf_{xx} + kf_{xy})^2 + (f_{xx}f_{yy} - f_{xy}^2)k^2$.
- What can we now conclude about the nature of the neighbourhood of the function at (a,b)?

Hessian matrix

- If $f_{xx} < 0$ and $f_{xx}f_{yy} - f_{xy}^2 > 0$ at (a, b) then $Q(0) < 0$ for all sufficiently small non-zero values of h and k, then f has a local maximum value at (a, b).
- If $f_{xx} > 0$ and $f_{xx}f_{yy} - f_{xy}^2 > 0$ at (a, b) then $Q(0) > 0$ for all sufficiently small non-zero values of h and k, then f has a local minimum value at (a, b).
- If $f_{xx}f_{yy} - f_{xy}^2 < 0$ at (a, b) there are combinations of small values for h and k for which $Q(0) > 0$ and other combinations of h and k for which $Q(0) < 0$. This means that f has a saddle point at (a, b).
- If $f_{xx}f_{yy} - f_{xy}^2 = 0$ at (a, b) another test is needed.

Hessian matrix

- The considerations of the previous slide show that determining whether there is a minimum or maximum at the point (a, b) boils down to looking at the following matrix and asking if it is positive-definite or not. This is the Hessian matrix.

$$\begin{bmatrix} f_{xx} & f_{xy} \\ f_{xy} & f_{yy} \end{bmatrix} \tag{2}$$

Hessian matrix

- A necessary and sufficient criterion of positive-definiteness for a Hermitian matrix (such as the Hessian matrix) is Sylvester's criterion - the determinant of every upper left $m \times m$ submatrix should be positive which means that $f_{xx} > 0$ and $f_{xx}f_{yy} - f_{xy}^2 > 0$ at (a, b). This condition corresponds to bullet point 2 on Slide 15.
- Note that the Hessian matrix is symmetric, so that in case of a local minimum it is a symmetric, positive-definite matrix which we know from the linear algebra part of this course is one that has positive eigenvalues.

Hessian Matrix

- For a local maximum we need to have a negative-definite matrix which means that $f_{xx} < 0$ and $f_{xx}f_{yy} - f_{xy}^2 > 0$ at (a, b). In this case, the determinant of every upper left $m \times m$ submatrix is negative if m is odd, and positive if m is even. The eigenvalues of the Hessian matrix are all negative in this case.
- The Hessian is the collection of all second-order partial derivatives. If f(x, y) is a twice (continuously) differentiable function, then
  $\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x}$ i.e., the order of differentiation does not matter, and the corresponding Hessian matrix is symmetric. The Hessian is denoted as $\nabla^2_{x,y} f(x, y)$

Example to find the absolute maximum and minimum in a closed bounded region

Find the absolute maximum and absolute minimum values of the function

$$f(x, y) = x^2 + y^2 - x - y$$

in the region $D = \{(x, y) : x^2 \leq y \leq 1\}$. (1

Sol: $f(x, y) = x^2 + y^2 - x - y$

Interior:

$f_x = 2x - 1$; $f_y = 2y - 1$

$f_x = f_y = 0 \Rightarrow$ Critical Point, (x, y)

$= \left(\frac{1}{2}, \frac{1}{2}\right)$

and it satisfies $x^2 \leq y$ and $y \leq 1$

Example (Ctd)

$$f_{xx} = 2, \quad f_{yy} = 2, \quad f_{xy} = 0$$

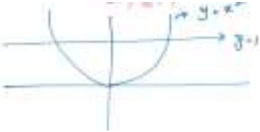
Since $f_{xx} > 0$, $f_{xx} f_{yy} - f_{xy}^2 = 4 > 0$,

f has relative minimum $-1/2$ at $(1/2, 1/2)$

Example (Ctd)

Boundary B_1 :

$$y = x^2, \quad -1 \le x \le 1$$



Let $g(x) = f(x, x^2)$

$$= x^4 - x$$

$g'(x) = 0 \implies 4x^3 - 1 = 0$

$\implies x = (1/4)^{1/3}$ and

$y = (1/4)^{2/3}$

Example(Ctd)

$g''(x) = 12x^2$

clearly, $g''(x) > 0$ at $x = (1/4)^{1/3}$

g has relative minimum at $x = (1/4)^{1/3}$ $\longrightarrow$

End Points : $x = -1$ and $x = 1$

$g(-1) = 2$ and $g(1) = 0$.

Example (Ctd)

On Boundary B_2:

$$y = 1, \quad -1 \le x \le 1$$

$$h(x) = f(x, 1) = x^2 - x$$

$$h'(x) = 0 \quad \Rightarrow \quad 2x - 1 = 0$$

$$\Rightarrow \quad x = 1/2 \quad \text{and} \quad y = 1.$$

$$h''(x) = 2 > 0.$$

h has relative minimum at $x = 1/2$

Example (Concluded)

End points : $x = -1$, $x = 1$.

Then, $h(-1) = 2$, $h(1) = 0$.

Absolute maximum at $(-1, 1)$. Value is 2

Absolute minimum at $(1/2, 1/2)$. Value is $-1/2$

Example to find local maximum and local minimum of a function

Find the local extreme values of the function

$$f(x,y) = xy - x^2 - y^2 - 2x - 2y + 4.$$

Solution

The function is defined and differentiable for all x and y. The

Function therefore has extreme values only at the points

Where f_x and f_y are simultaneously zero. This leads to

$$f_x = y - 2x - 2 = 0 \qquad f_y = x - 2y - 2 = 0,$$

Or

$$x = y = -2.$$

Therefore the point (2,-2) is the only point where f may take on a extreme value. Therefore,

$$f_{xx} = -2, \qquad f_{yy} = -2, \qquad f_{xy} = 1.$$

Solution(Ctd)

The discriminant of f at (a,b) = (-2,-2) is

$f_{xx}f_{yy} - f_{xy}^2 = (-2)(-2) - (1)^2 = 4 - 1 = 3.$

The combination

$f_{xx} < 0$ and $f_{xx}f_{yy} - f_{xy}^2 > 0$

Tells us that f has a local maximum at (-2,-2) . The value Of f at this point is $f(-2,-2) = 8$.