# ĐỒ ÁN MÔN HỌC

## Abstract

## Introduction

## Optimize Algorithm

1. Gradient Descent
   1. Ý tưởng chính của Gradient Descent

Trong lĩnh vực Machine Learning và Toán Tối ưu nói chung, việc tìm giá trị nhỏ nhất (hoặc lớn nhất của hàm số) là rất quan trọng . Gradient Descent là một trong những thuật toán tối ưu quan trọng và phổ biến nhất trong lĩnh vực *Machine Learning* và *Deep Learning.* Mục tiêu quan trọng của Gradient Descent là tìm giá trị tham số tối ưu của hàm mất mát sao cho hàm mất mát đạt giá trị nhỏ nhất.

Ta gọi:

Để hàm mất mát đạt cực tiểu tại điểm , độ dốc dạo hàm của hàm mất mát bằng không:

Tuy nhiên, hàm số trong thực tế rất phức tạp, phi tuyến, nhiều chiều, thường rất khó để tìm ra nghiệm tối ưu nhất cho bài toán. Do đó, ý tưởng chính để giải quyết bài toán này là tìm một phương pháp xấp xỉ hàm - sử dụng những phương trình đơn giản, dễ đạo hàm với giả định rằng sau một thời gian *t* cập nhật giá trị, điểm sẽ xấp xỉ ).

Để xem xét về hướng đi của bước để đưa x về điểm tối ưu, ta khai triển chuỗi Taylor bậc 1 xấp xỉ hàm xung quanh điểm , ta có được:

*với mọi nằm gần*

Sau khi đã xấp xỉ bằng biểu thức khai triển Taylor trên, ý tưởng chính của Gradient Descent là di chuyển sao cho hàm được tối thiểu hóa. Dựa vào biểu thức xấp xỉ trên, bằng cách di chuyển bằng một đại lượng > 0 theo hướng ngược hướng với đạo hàm , ta biểu diễn được bằng công thức sau:

Ta gọi là điểm khởi tạo ban đầu, tham số được xác địng là bước nhảy *(learning rate)*

* 1. *L-smooth* functions (hàm *L-smooth*)

Để có thể đưa ra được tốc độ hội tụ cụ thể cho thuật toán Gradient Descent, ta đặt ra giới hạn (bound) về mức độ thay đổi của gradient khi ta di chuyển một chút theo bất kỳ hướng nào. Khi ta bắt đầu di chuyển điểm theo một hướng ngược hướng với đạo hàm từ , giá trị gradient mới sẽ sẽ không thay đổi quá nhanh so với , từ đó giữ cho luôn là một đại lượng độ dốc giúp hàm hội tụ ổn định trong 1 khoảng thời gian

Giả sử, độ dốc đạo hàm là liên tục 𝐿-Lipschitz với hằng số , điều kiện này trên lý thuyết thường được gọi là *L-smoothness*. Xét hàm khả vi trong miền lồi . Hàm được gọi là *L-smoothness* nếu đạo hàm của nó là liên tục 𝐿-Lipschitz, ta có:

(1)

Hệ quả quan trọng của *L-smoothness* được phát biểu rằng: giá trị của hàm có thể bị thay đổi, và bị chặn trên bởi một biểu thức bậc 2:

Biểu thức trên có thể được chứng minh như sau:

Với bất đẳng thức Cauchy-Schwarz, ta có:

Xét khoảng cách giữa điểm hiện tại so với x, sau khi qua thời điểm:

Từ định nghĩa (1), ta có:

Thay vào (2), ta có:

Để chứng minh mỗi bước Gradient Descent đều làm giảm giá trị hàm mục tiêu , ta xét điểm cập nhật tiếp theo với bước cập nhật . Từ , ta có:

Chọn bước nhảy

từ ta có:

Ta thấy nếu là *L-smooth*, với giá trị , tại mỗi bước của Gradient Descent luôn đảm bảo gradient của hàm số luôn giảm với ít nhất một lượng tỉ lệ với bình phương độ lớn gradient nếu gradient của hàm số hiện tại khác 0.

### SGD

### Adagrad

### Momentum

### RMSprops

### Adam

### AdamW

## Experiment