1. [Không bắt buộc] Nêu các vấn đề của Gradient descent, lý do và cách khắc phục.

Local minima: không tìm được cực trị của Cost function vì bị kẹt ở local minina và plateus.

Feature scales: khi giữa các feature có giá trị quá chênh lệch nhau thì việc tìm cực trị của cost function mất rất nhiều thời gian.

1. So sánh Batch gradient decent và Stochastic gradient decent (điểm giống nhau, khác nhau, ưu điểm, khuyết điểm).

Khác nhau: BGDdùng toàn bộ dử liệu để tính toán mỗi khi thử một theta mới, SGD chỉ dùng 1 sample.

Giống nhau: Đều thử đi thử lại để tìm theta làm cho cost function có giá trị nhỏ nhất.

Ưu điểm: BGD có thể chạy tốt trên tập dử liệu có nhiều feature, cost fucntion luôn giảm sau mỗi lần thử. SGD có thể vừa chạy tốt trên tập dử liệu có nhiều feature vừa có nhiều sample, có thể vượt qua local minima.

Khuyết điểm: BGD chạy chậm khi tạp dử liệu có quá nhiều sample. SGD cần nhiều vòng lặp để tìm ra cực trị của cost function.

1. Trình bày cách huấn luyện các Polynomial regression models.

Khi trainning Polynomial regression models ta cần tạo ra feature cho các theta có bậc cao bằng cách nhân các feature bậc thấp tương ứng. Sau đó, ta có thể dùng Linear regression để tranning như bình thường.

1. Nêu đặc điểm learning curves của underfitting, overfitting và good-fitting models.

Underfitting: khi model quá đơn giản so với dử liệu của bài toán, learning curve có xu hướng đi xuống.

Overfitting: khi model quá phức tạp so với dử liệu của bài toán, learning curve có xu hướng đi xuống, nhưng có sai số cao hơn.

Good-fitting: khi model phù hợp với dử liệu của bài toán, learning curve có xu hướng đi ngang và có sai số thấp.