1. Giả sử khi dùng Batch gradient descent và vẽ learning curve của nó (đường validation error theo epoch), ta thấy validation error chỉ tăng lên chứ không giảm. Nguyên nhân của hiện tượng này có thể là gì? Làm sao khắc phục?

Nguyên nhân của hiện tượng này là overfitting do trainning quá lâu. Để khắc phục ta cần dùng Early Stopping, Regularization.

1. Để tránh overfitting của Stochastic gradient descent do train quá lâu ta có thể sử dụng phương pháp nào? Mô tả cách cài đặt phương pháp này cho SGD.

Hiệu tượng overfitting do trainning quá lâu đã xãy ra. Ta có thể sử dụng kỹ thuật Early Stopping để tránh việc này.

Để cài đặt ta dựa vào hàm MSE, khi hàm này bắt đầu tăng theo sau mỗi lần trainning thì ta dừng tranning và lấy model trước đó làm kết quả. Tuy nhiên, MSE cùa Stochastic gradient descent có thể có các cực trị, nên ta cần chạy thêm khoảng 200 (chọn tùy ý) để chắc chắn MSE sẻ tiếp tục tăng, nếu không thì ta cần tiếp tục trainning.

1. Khi huấn luyện một Polynomial regression ta thấy learning curves theo training error và validation error có khoảng cách lớn. Hiện tượng gì đã xảy ra? Nêu các cách khắc phục.

Hiện tượng OverFitting đã xãy ra, trên tập trainning thì chạy đúng còn trên tập test thì sai. Để khắc phục ta cần cho thêm dử liệu, thực hiện regularization.

1. Khi huấn luyện một Ridge regression model ta thấy training error và validation error gần như bằng nhau và có giá trị khá cao. Model của chúng ta đã bị vấn đề gì? Để giải quyết ta nên tăng hay giảm hệ số regularization ?

Training error và validation error gần như bằng nhau và có giá trị khá cao thì model đã bị underfitting, vì cả trên tập test và train đề chạy không đúng. Để giải quyết ta chỉ cần tăng hệ số .

1. Tại sao ta nên dùng Ridge regression, hoặc các regularization models khác, hơn là dùng Linear regression hoặc Polinomial regression?

Vì dùng Linear regression hoặc Polinomial regression thì dể dẫn đến underfitting làm cho dử liệu không được mô tả chính xác. Trong khi, dùng Ridge regression, hoặc các regularization models khác ta có thể dùng các model bậc cao và tinh chỉnh lại các tham số để có được good fitting.

1. So sánh Ridge regression, Lasso regression, và Elastic net: điểm giống nhau, khác nhau, ưu điểm, khuyết điểm.

Giống nhau: Đều thuộc loại thuật toán Regularization dùng để giảm giá trị của theta.

Khác nhau: Lasso regression làm giảm tham số theta nhanh hơn Ridge regression, trong khi đó Elastic net kết hợp cả hai và dùng tham số r là tỷ lệ của hai tham số a trong hai thuật toán trên