1. Trình bày ý tưởng của regularization. Tại sao regularization có thể giúp ngăn overfitting?

Regularization là phương pháp làm giảm giá trị của theta từ đó làm giảm ảnh hưởng của các feature bậc cao trong model. Khi giảm theta ta củng làm giảm giá trị của các cực trị trong model, vì vậy khi ta dùng model quá phức tạp so với dử liệu (overfitting) thì model cũng không có sai lệch quá lớn so với dử liệu.

1. Nếu cần train 1 data set có một triệu features bạn sẽ chọn thuật toán training nào: Normal equation hay Gradient descent? Tại sao?

Vì Normal equation cần phải tìm ma trận nghịch đảo của ma trận số cột là số các feature, nên nếu số feature quá lớn thì dùng Normal equation sẻ không phù hợp. Trong trường hợp này nên chọn Gradient descent.

1. Nếu data set của bạn có các features với khoảng giá trị rất khác nhau thì thuật toán training nào (trong 2 thuật toán: Normal equation hay Gradient descent) sẽ gặp vấn đề? Vấn đề đó là gì? Làm sao để xử lý?

Gradient descent sẻ bị ảnh hưởng. Đó là feature scalling. Khi đó ta cần chia feature có giá trị cao cho một con số nào đó để có giá trị gần bằng các feature khác.

1. Nếu dùng Gradient descent để train một logistic regression model thì nó có thể bị mắc kẹt ở local minima không? Tại sao?

Vì cost function của logistic regression model không có minima nên Nếu dùng Gradient descent thì sẻ không bị mắc kẹt.

1. Nếu chạy các thuật toán Gradient descent đủ lâu, chúng có đưa đến các hypothesis giống nhau không? Tại sao?

Khi chạy các thuật toán Gradient descent quá lâu thì tất các thuật toán đều cho ra hypothesis bị over fitting, nên các hypothesis function này sẻ giống nhau.