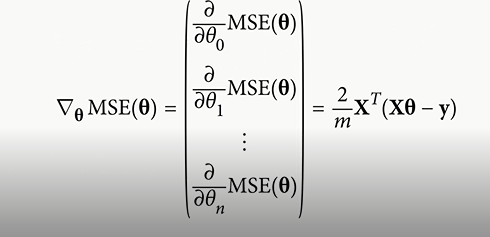
1. Các vấn đề của Gradient descent:

Local minima (kẹt ở cực trị địa phương): khi cost function phức tạp. Cách khắc phục: khởi tạo theta hợp lý.

Feature scales: khi feature khác scales.

Cách khắc phục: trước khi đưa dữ liệu vào thì phải feature scaling dùng standard scaler.

1. Batch gradient decent:

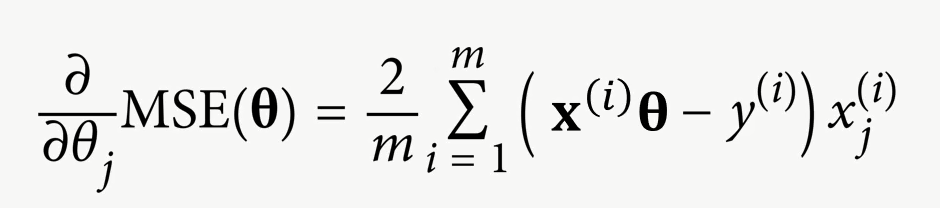


\* Khác nhau: m là số lượng sample

\* Ưu điểm: làm nhanh hơn Normal Equation, chính xác hơn SGĐ, chạy ít vòng lặp hơn SGD

\* Nhược điểm: tốc độ tính đạo hàm còn chậm khi tập training set lớn, dùng quá nhiều memory, đôi khi không chạy luôn

Stochastic gradient descent (SGD):



\* Khác nhau: chỉ dùng 1 sample ngẫu nhiên

\* Ưu điểm: chạy nhanh hơn nhiều lần so với BGD và Normal Equation, sử dụng ít memory hơn BGD, có thể bỏ qua hiện tượng Local minima

\* Nhược điểm: chạy không chính xác bằng BGD, đôi khi chỉ hướng đi sai, chạy nhiều vòng lặp hơn, gần tới minimum thì tiếp tục bounce around (khắc phục bằng cách cho số vòng lặp hoặc cho learning rate là hàm giảm dần (không quá nhanh hoặc quá chậm) theo thời gian)

\*Giống nhau: công thức tính, chạy bằng cách tính đạo hàm.