Fullname: Nguyễn Ngọc Thiện

Student ID: 19110148

Class: Machine Learning\_ Nhom 02CLC

1. Nêu các vấn đề của Gradient descent, lý do và cách khắc phục.

Gradient descent có 2 vấn đề như sau:

* Local minimal: hàm số đạt cực tiểu tại vị trí local (đạo hàm bằng 0) chứ không đạt được tại vị trí global. Do điều kiện dừng bằng không
* Feature scales: do giá trị giữa các feature quá khác biệt dẫn đến việc tìm ra global mininal chậm (đi vòng, tốn nhiều vòng lập) và thậm chí là không chính xác.

Giải pháp:

|  |  |
| --- | --- |
| Local minimal | Feature scales |
| Sử dụng các thuật toán local search: random restart để thực hiện ramdom nhiều lần cho đến khi tìm được giá trị nhỏ nhất (cực tiểu), cần phải giới hạn vòng lập.  Simulated annealing: thực hiện “rung” để giúp cho model thoát khỏi các đoạn bằng phẳng (plateau). | Xử lí các giá trị khác biệt  Giới hạn các vòng lập vì để giá trị = 0 thì khi tới global nó có thể lập khá nhiều lần. |

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

Giống nhau: đều dùng đạo hàm để tính theta và cập nhật theta qua mỗi lần lập, tìm bộ tham số cực tiểu.

|  |  |  |
| --- | --- | --- |
|  | Batch gradient decent | Stochastic gradient decent |
| Khác nhau | Dùng nhiều dữ liệu nhất để tính toán các đạo hàm (tất cả dữ liệu) | Dùng ít dữ liệu nhất có thể để tính toán đạo hàm (1 sample) |
| Ưu điểm | Đối với nhiều feature thì tốc độ vẫn chạy nhanh, cost function có sai số nhỏ hơn và luôn tăng  Ít lần lập trên 1 lần chạy | Nhanh hơn BGD vì nó chỉ thực hiện đạm hàm trên 1 sample. Vẫn chạy nhanh trên tập dữ liệu có nhiều sample và nhiều feature  Giúp vượt qua local minimal |
| Khuyết điểm | Chạy chậm hơn SGD, gặp trở ngại trên tập dữ liệu nhiều sample vì nó thực hiện đạo hàm qua mỗi sample. | Cost function có sai số lớn và biến động tăng giảm thất thường. Dùng nhiều vòng lập hơn và dễ xảy ra tình trạng ‘chạy liên tục xung quanh cực tiểu’ vì giá trị theta là xấp xỉ bằng 0 |
| Áp dụng | Ít được sử dụng hơn vì khá tốn thời gian và phần lớn tập dữ liệu cần train khá lớn | Được sử dụng nhiều vì nó nhanh hơn và cũng tìm ra được cực tiểu. |

1. Trình bày cách huấn luyện các Polynomial regression models.
   * Giống như linear regression ta cũng cần phải tính hypothesis dùng để mô tả dữ liệu, đối với Polynomial regression ta cần thêm các feature bậc cao (bậc 2, 3, 4,...). Tuy nhiên bậc càng cao sẽ xảy ra overfitting và ngược lại và underfitting.
   * Tính MSE để tìm ra đâu là đường mô tả hiệu quả nhất (sử dụng NE hoặc BGD) tùy thuộc vào sample và feature mà ta có thể quyết định sử dụng phương thức nào
   * Chạy Linear regression để huấn luyện thuật toán (e.g., Normal Eq., Gradient descent).
2. Nêu đặc điểm learning curves của underfitting, overfitting và good-fitting models

Điểm chung là có MSE rất cao khi ít sample và xu hướng ổn định khi dữ liệu lớn

**Đặc điểm**

* + Underfitting: bài toán quá đơn giản khi dùng mô tả dữ liệu, không thể khái quả được hết dữ liệu. Learning curves có xu hướng giảm khi tập dữ liệu lớn lên. Gap của curves khá là gần nhau và MSE khá cao xấp xĩ 1.7
  + Overfitting: bài toán quá phức tạp, cố gắng đi qua hầu hếu các điểm để mô tả dữ liệu, tuy nhiên sẽ gặp khó khăn khi có quá nhiều điểm dữ liệu nằm gần nhau. Learning curves cũng có xu hướng giảm khi tập dữ liệu lớn lên. Tuy nhiên Gap của curves khá lớn -> sai số cao. Có MSE tương đối xấp xĩ 1.2
  + Good-fitting: Learning curves cũng có xu hướng giảm khi tập dữ liệu lớn lên. Gap của curves khá là gần nhau và MSE khá cao xấp xĩ 1. Model khác tốt, tuy nhiên trên thực tế rất hiếm gặp trường hợp này.

**Demo Program (double click to open)**

****