

|  |  |
| --- | --- |
| Hours(x) | Points(y) |
| 3 | 4 |
| 4 | 5 |
| 5 | 6 |
| 6 | 7 |

Đây là ví dụ em cho để chạy tay

Cho W = 0 , ta có hàm loss = (y^-y)^2 = (w\*x – y)^2 như bảng sau

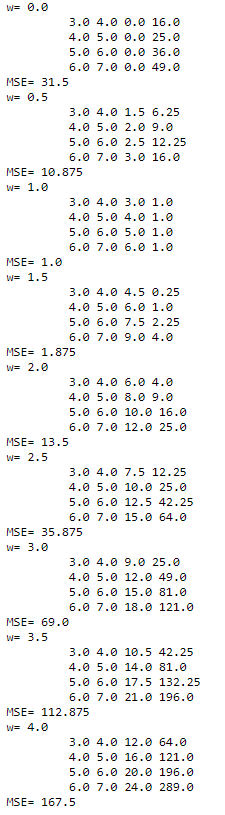
|  |  |  |  |
| --- | --- | --- | --- |
| Hours(x) | Poin(y) | y^ = w\*x(w = 0) | Loss |
| 3 | 4 | 0 | 16 |
| 4 | 5 | 0 | 25 |
| 5 | 6 | 0 | 36 |
| 6 | 7 | 0 | 49 |
|  |  |  | MSE =126/4=31.5 |

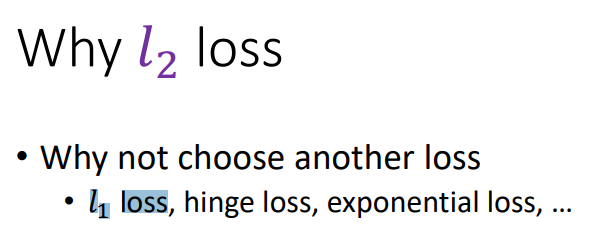
Tương tự ta có :

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Hours(x) | Poin(y) | Loss(w=0) | Loss(w=1) | Loss(2) | Loss(3) | Loss(4) |
| 3 | 4 | 16 | 1 | 4 | 25 | 64 |
| 4 | 5 | 25 | 1 | 9 | 49 | 121 |
| 5 | 6 | 36 | 1 | 16 | 81 | 196 |
| 6 | 7 | 49 | 1 | 25 | 121 | 289 |
|  |  | MSE=31.5 | MSE=1 | MSE=13.5 | MSE=69 | MSE=167.5 |

Như thế ta thấy điểm có giá trị nhỏ nhất của loss là tại W = 1

Đây là kết quả chạy được :





Bởi vì là hàm bình phương có đạo hàm tại mọi nơi, trong khi hàm trị tuyệt đối thì không (đạo hàm không xác định tại 0)

*nghiệm cuối cùng của Gradient Descent phụ thuộc rất nhiều vào điểm khởi tạo và learning rate.*