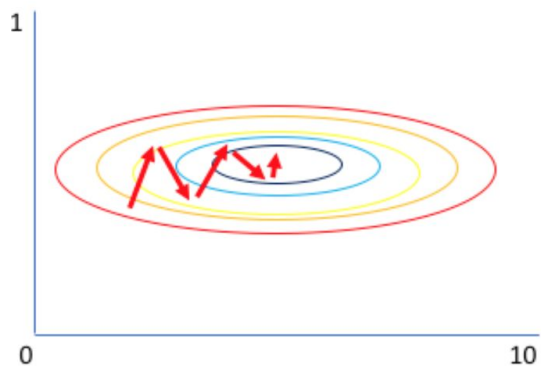


Normalization

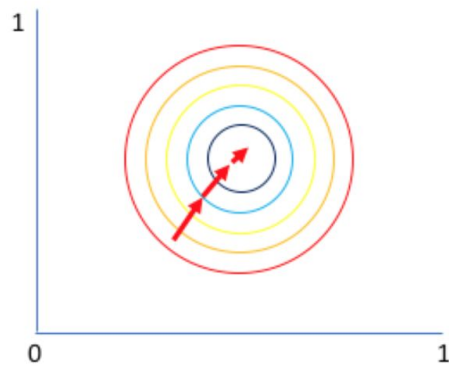
Problem

```
1 import numpy as np
2
3 x = np.array([[1.5, 2864, 2.3],
4              [2.6, 8372, 1.8],
5              [1.2, 6453, 2.2],
6              [2.3, 9587, 3.7],
7              [1.9, 2332, 3.1],
8              [3.7, 8574, 1.5],
9              [2.1, 7665, 2.3],
10             [1.4, 2428, 1.8],
11             [3.7, 9476, 3.2],
12             [1.5, 3422, 2.4]])
```

Why?



Gradient of larger parameter
dominates the update




Both parameters can be
updated in equal proportions

Normalization

- adjusting values measured on different scales to a notionally common scale

x_1	x_2	x_3
0.00234	387428	23.53
0.00129	43223	76.05
0.00943	234004	15.43
0.01202	48329	9.93



x_1	x_2	x_3
0.2	0.9	0.2
0.1	0.2	0.8
0.7	0.7	0.3
0.9	0.3	0.1

Normalization

- Min Max
$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

- Standard Score

$$z = \frac{x - \mu}{\sigma}$$

μ = Mean

σ = Standard Deviation

```
from sklearn.preprocessing import MinMaxScaler, StandardScaler
X = [[1,2],[2,1],[3,2],[4,3]]
scaler = MinMaxScaler()
scaler.fit(X)
X = scaler.transform(X)
print(X)
```

Example

Lab 11

Normalization

Lab 11

- Normalize **Auto MPG data**.
- Normalize it using `MinMax`.
- Use sklearn package in Python
- You shouldn't normalize MPG column.
- Submit on Blackboard.

Auto MPG

- For example,
 - <https://archive.ics.uci.edu/ml/index.php>
 - <https://archive.ics.uci.edu/ml/datasets/Auto+MPG>
- This data is to predict MPG given 7 features of a car (excluding 'car name')
 - mpg: continuous
 - cylinders: multi-valued discrete
 - displacement: continuous
 - horsepower: continuous
 - weight: continuous
 - acceleration: continuous
 - model year: multi-valued discrete
 - origin: multi-valued discrete
 - car name: string (unique for each instance)

Auto MPG

18.0	8	307.0	130.0	3504.	12.0	70	1	"chevrolet chevelle malibu"
15.0	8	350.0	165.0	3693.	11.5	70	1	"buick skylark 320"
18.0	8	318.0	150.0	3436.	11.0	70	1	"plymouth satellite"
16.0	8	304.0	150.0	3433.	12.0	70	1	"amc rebel sst"
17.0	8	302.0	140.0	3449.	10.5	70	1	"ford torino"
15.0	8	429.0	198.0	4341.	10.0	70	1	"ford galaxie 500"
14.0	8	454.0	220.0	4354.	9.0	70	1	"chevrolet impala"
14.0	8	440.0	215.0	4312.	8.5	70	1	"plymouth fury iii"
14.0	8	455.0	225.0	4425.	10.0	70	1	"pontiac catalina"
15.0	8	390.0	190.0	3850.	8.5	70	1	"amc ambassador dpl"
15.0	8	383.0	170.0	3563.	10.0	70	1	"dodge challenger se"