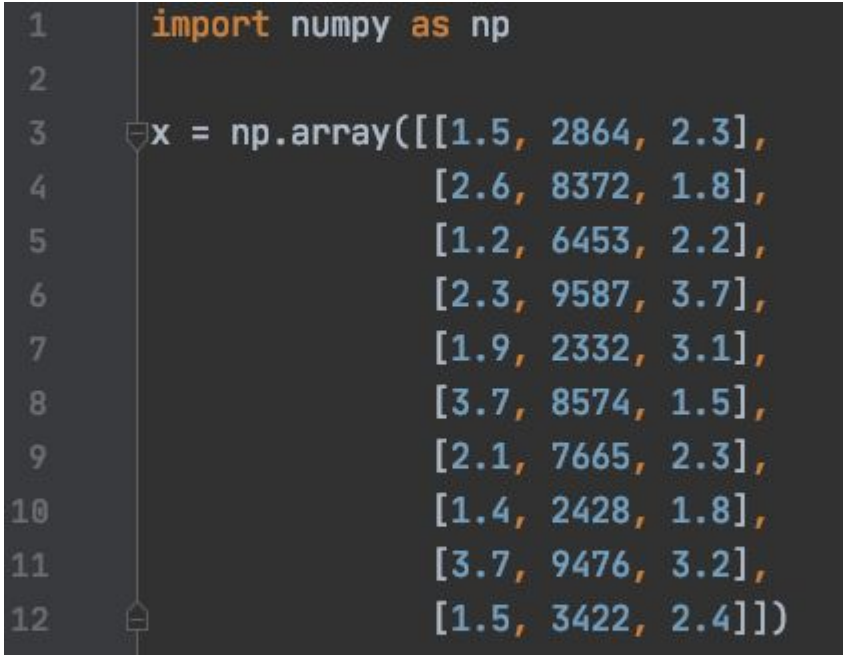
Normalization

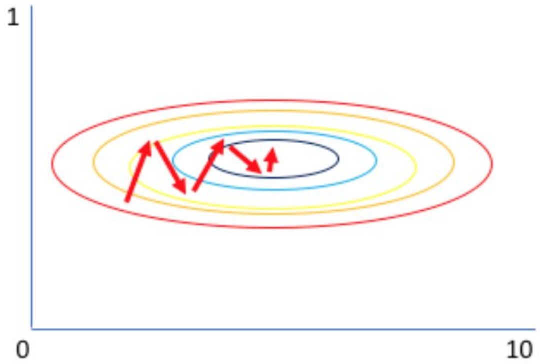
Problem



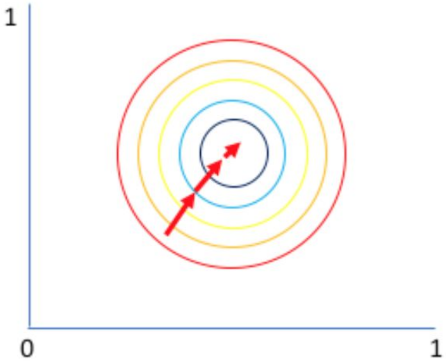
```python
import numpy as np

x = np.array([[1.5, 2864, 2.3],
              [2.6, 8372, 1.8],
              [1.2, 6453, 2.2],
              [2.3, 9587, 3.7],
              [1.9, 2332, 3.1],
              [3.7, 8574, 1.5],
              [2.1, 7665, 2.3],
              [1.4, 2428, 1.8],
              [3.7, 9476, 3.2],
              [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

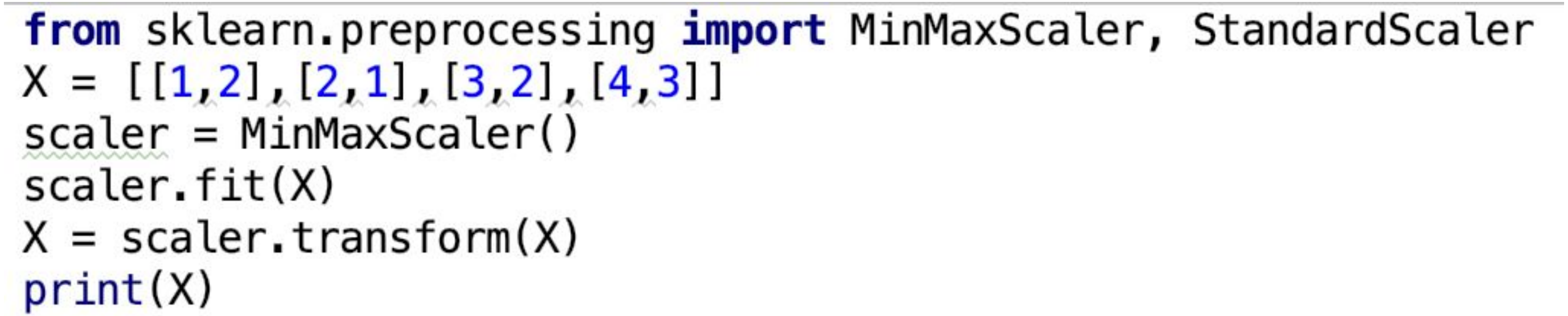
$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

- Standard Score

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

$\mu =$ Mean

$\sigma =$ 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