



Scaling to vector unit norm


Scaling to unit norm

In scaling to unit norm, we divide each **feature vector** by either the Manhattan distance (l1 norm) or the Euclidean distance of the vector (l2 norm).


$$l1(X) = |x1| + |x2| + \dots + |xn|$$

$$l2(X) = \text{sqr}(x1^2 + x2^2 + \dots + xn^2)$$

Scaling to unit norm: across features

| Price | | Price |
|-------|---|-------|
| 100 |  | 0.73 |
| 90 | | 0.55 |
| 50 | | -0.18 |
| 40 | | -0.36 |
| 20 | | -0.73 |
| 100 | | 0.73 |
| 50 | | -0.18 |
| 60 | | 0.00 |
| 120 | | 1.09 |
| 40 | | -0.36 |
| 200 | | 2.55 |


Scaling to unit norm: across features



| Price | Price |
|-------|-------|
| 100 | 0.73 |
| 90 | 0.55 |
| 50 | -0.18 |
| 40 | -0.36 |
| 20 | -0.73 |
| 100 | 0.73 |
| 50 | -0.18 |
| 60 | 0.00 |
| 120 | 1.09 |
| 40 | -0.36 |
| 200 | 2.55 |

Scaling to unit norm: across features

| Price | Price |
|-------|-------|
| 100 | 0.73 |
| 90 | 0.55 |
| 50 | -0.18 |
| 40 | -0.36 |
| 20 | -0.73 |
| 100 | 0.73 |
| 50 | -0.18 |
| 60 | 0.00 |
| 120 | 1.09 |
| 40 | -0.36 |
| 200 | 2.55 |

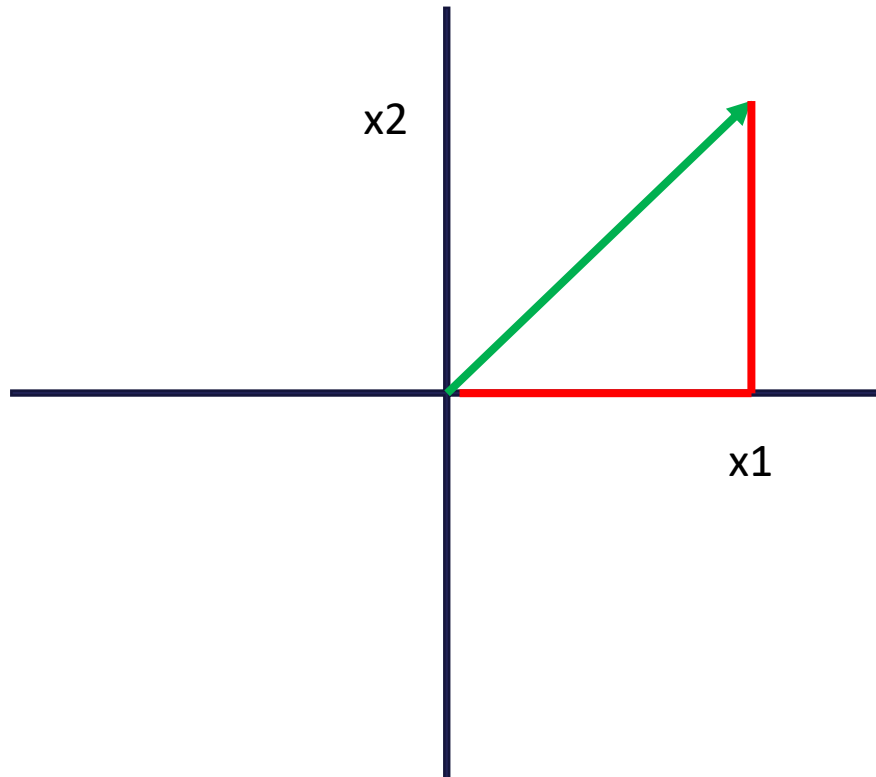


| | Gender | Price | Make | Engine |
|--|--------|-------|------|--------|
| | 1 | 100 | 1 | 2000 |
| | 0 | 90 | 1 | 2000 |
| | 0 | 50 | 2 | 1500 |
| | 0 | 60 | 2 | 2200 |
| | 1 | 3 | 3 | 3000 |
| | 1 | 120 | 4 | 4500 |
| | 0 | 200 | 4 | 4500 |

Scaling to unit norm: example

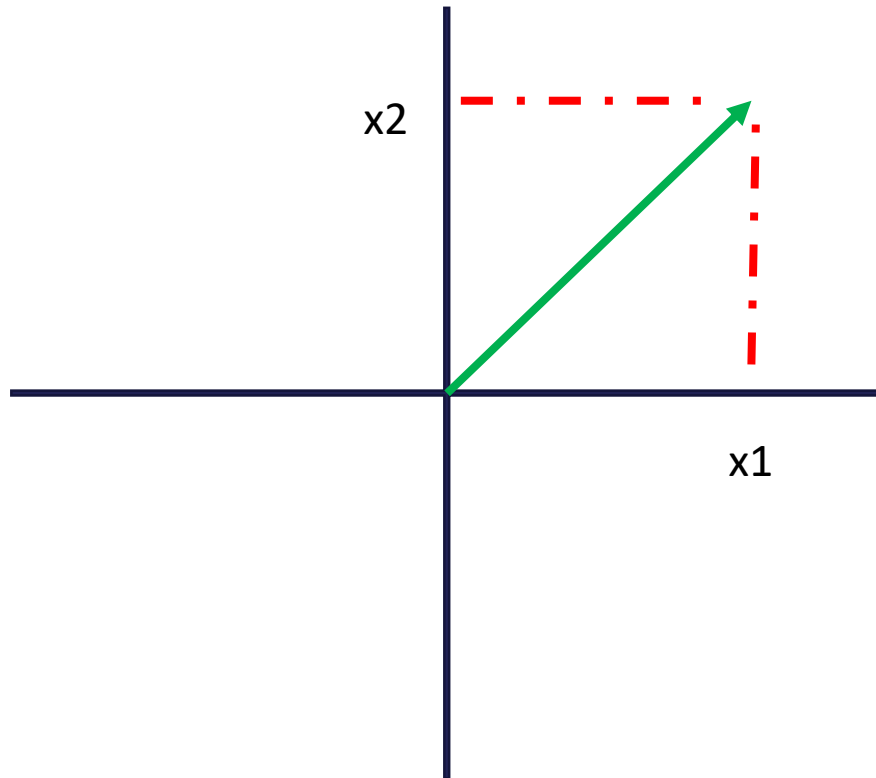
| Gender | Price | Make | Engine | | | L2 Norm | | | Gender | Price | Make | Engine |
|--------|-------|------|--------|--------------------------|--|---------|-------------------------|--|--------|-------|-------|--------|
| 1 | 100 | 1 | 2000 | Norm of each vector ➔ | | 2002 | Divide by the norm ➔ | | 0.000 | 0.050 | 0.000 | 0.999 |
| 0 | 90 | 1 | 2000 | | | 2002 | | | 0.000 | 0.045 | 0.000 | 0.999 |
| 0 | 50 | 2 | 1500 | | | 1501 | | | 0.000 | 0.033 | 0.001 | 0.999 |
| 0 | 60 | 2 | 2200 | | | 2201 | | | 0.000 | 0.027 | 0.001 | 1.000 |
| 1 | 3 | 3 | 3000 | | | 3000 | | | 0.000 | 0.001 | 0.001 | 1.000 |
| 1 | 120 | 4 | 4500 | | | 4502 | | | 0.000 | 0.027 | 0.001 | 1.000 |
| 0 | 200 | 4 | 4500 | | | 4504 | | | 0.000 | 0.044 | 0.001 | 0.999 |

Manhattan distance, l1



$$L1 = |x1| + |x2|$$

Euclidean distance, l2



$$L2 = \sqrt{x_1^2 + x_2^2}$$



• Normalisation to unit length

- Clustering
- Text Analytics



Accompanying Jupyter Notebook



- Read the accompanying Jupyter Notebook
- Scaling to unit norm with Scikit-learn

THANK YOU

www.trainindata.com