Train In Data

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).

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

$$l2(X) = sqr(x1^2 + x2^2 + \ldots + xn^2)$$

Scaling to unit norm: across features

Price
100
90
50
40
20
100
50
60
120
40
200

Price
0.73
0.55
-0.18
-0.36
-0.73
0.73
-0.18
0.00
1.09
-0.36
2.55

Scaling to unit norm: across features

Price
100
90
50
40
20
100
50
60
120
40
200

Price
0.73
55
0.18
-0.36
-0.73
0.73
-0.18
00
1.09
-0.36
2.55

Scaling to unit norm: across features

Price
100
90
50
40
20
100
50
60
120
40
200

Price
0.73
55
0.18
-0.36
-0.73
0.73
-0.18
00
1.09
-0.36
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
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

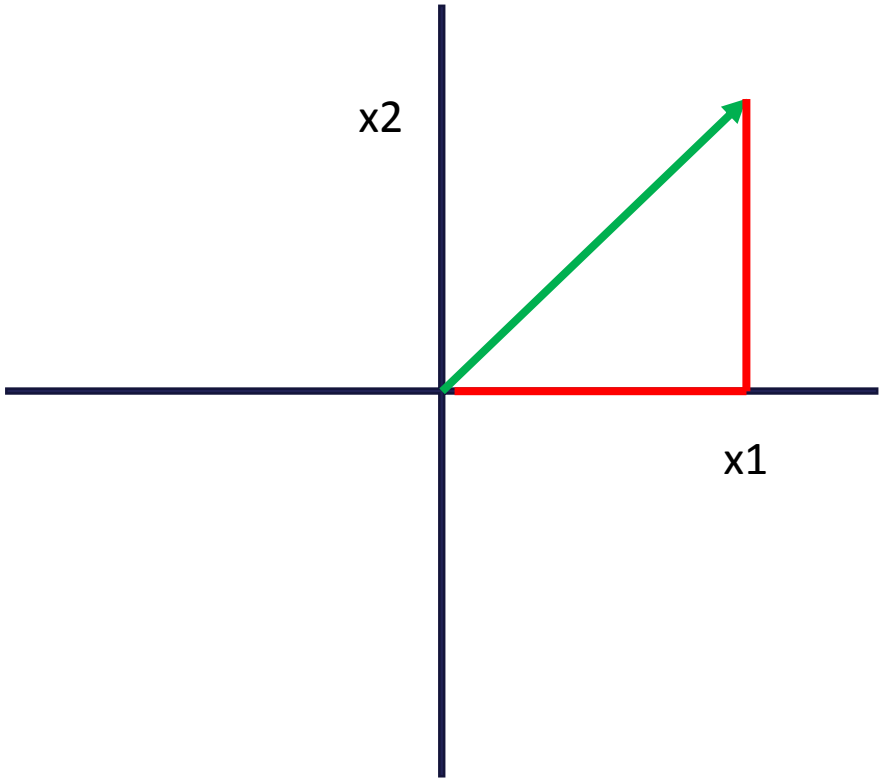
Norm of each vector

L2 Norm
2002
2002
1501
2201
3000
4502
4504

Divide by the norm

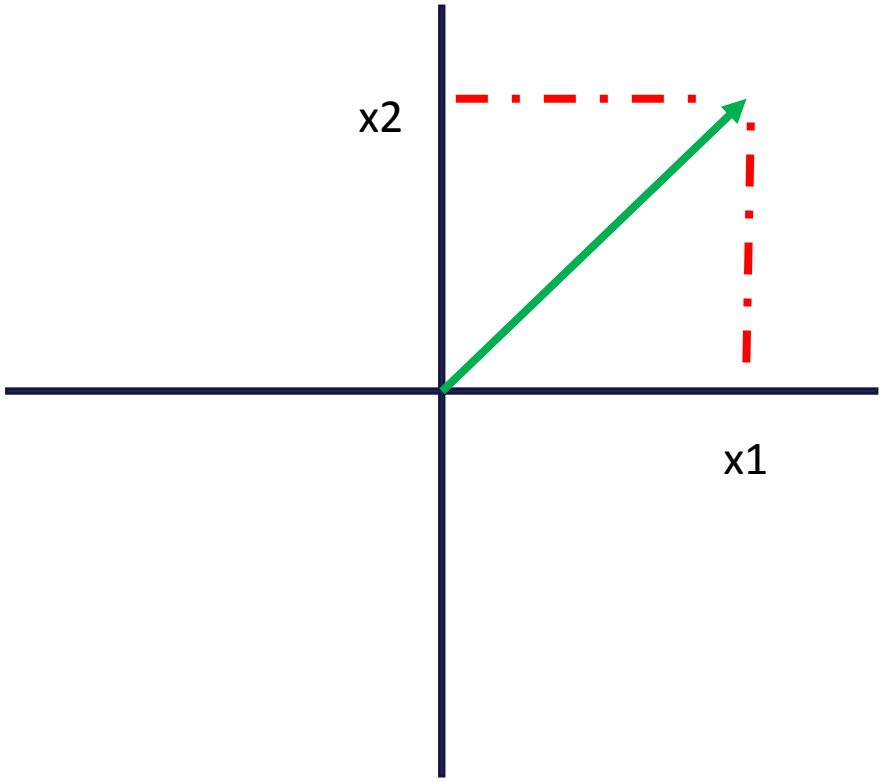
Gender	Price	Make	Engine
0.000	0.050	0.000	0.999
0.000	0.045	0.000	0.999
0.000	0.033	0.001	0.999
0.000	0.027	0.001	1.000
0.000	0.001	0.001	1.000
0.000	0.027	0.001	1.000
0.000	0.044	0.001	0.999

Manhattan distance, l1



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

Euclidean distance, l2



$$L2 = \sqrt{x1^2 + x2^2}$$

Normalisation to unit length

- Clustering
- Text Analytics

Accompanying Jupyter Notebook

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

Train In Data

THANK YOU

www.trainindata.com