KNN for Classification

- KNN for Classification

	history	*science*	*research*	*offers*	*students*	*hall*	Class
Anthropology	0	0.537	0.477	0	0.673	0.177	A
Art	0	0	0	0.961	0.195	0.196	B
Biology	0	0.347	0.924	0	0.111	0.112	A
Chemistry	0	0.975	0	0	0.155	0.158	A
Communication	0	0	0	0.780	0.626	0	B
Computer Science	0	0.989	0	0	0.130	0.067	A
Criminal Justice	0	0	0	0	1	0	B
Economics	0	0	1	0	0	0	A
English	0	0	0	0.980	0	0.199	B
Geography	0	0.849	0	0	0.528	0	A
History	0.991	0	0	0.135	0	0	B
Mathematics	0	0.616	0.549	0.490	0.198	0.201	A
Modern Languages	0	0	0	0.928	0	0.373	B
Music	0.970	0	0	0	0.170	0.172	B
Philosophy	0.741	0	0	0.658	0	0.136	B
Physics	0	0	0.894	0	0.315	0.318	A
Political Science	0	0.933	0.348	0	0.062	0.063	A
Psychology	0	0	0.852	0.387	0.313	0.162	A
Sociology	0	0	0.639	0.570	0.459	0.237	A
Theatre	0	0	0	0	0.967	0.254	? (B)

```
Cos(x, y) = x . y / ||x|| * ||y||
```

x . y = product (dot) of the vectors ‘x’ and ‘y’.
||x|| and **||y||** = length of the two vectors ‘x’ and ‘y’.
||x|| * ||y|| = cross product of the two vectors ‘x’ and ‘y’.

Consider an example to find the similarity between two vectors – **'x'** and **'y'**, using Cosine Similarity.
The 'x' vector has values, **x = { 3, 2, 0, 5 }**
The 'y' vector has values, **y = { 1, 0, 0, 0 }**

```
x . y = 3*1 + 2*0 + 0*0 + 5*0 = 3

||x|| = √ (3)^2 + (2)^2 + (0)^2 + (5)^2 = 6.16

||y|| = √ (1)^2 + (0)^2 + (0)^2 + (0)^2 = 1

∴ Cos(x, y) = 3 / (6.16 * 1) = 0.49
```

KNN for Classification

- KNN for Classification

Document	Class	Similarity to Theatre
Criminal Justice	B	0.967075
Anthropology	A	0.695979
Communication	B	0.605667
Geography	A	0.510589
Sociology	A	0.504672
Physics	A	0.385508
Psychology	A	0.343685
Mathematics	A	0.242155
Art	B	0.238108
Music	B	0.207746
Chemistry	A	0.189681
Computer Science	A	0.142313
Biology	A	0.136097
Modern Languages	B	0.0950206
Political Science	A	0.0762211
English	B	0.0507843
Philosophy	B	0.0345299
History	B	0
Economics	A	0

KNN for Classification

- KNN for Classification
  - 1-NN: B
  - k-NN:
    - 3-NN: B
    - 5-NN: A

KNN for Regression

- KNN for Regression

Age	Loan	House Price Index	Distance	
25	$40,000	135	102000	
35	$60,000	256	82000	
45	$80,000	231	62000	
20	$20,000	267	122000	
35	$120,000	139	22000	2
52	$18,000	150	124000	
23	$95,000	127	47000	
40	$62,000	216	80000	
60	$100,000	139	42000	3
48	$220,000	250	78000	
33	$150,000	264	8000	1
48	**$142,000**	**?**		

$$D = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2}$$

KNN for Regression

- We can now use the training set to classify an unknown case (Age=48 and Loan=$142,000) using Euclidean distance.
- If K=1 then the nearest neighbour is the last case in the training set with HPI=264.
- D = Sqrt[(48-33)^2 + (142000-150000)^2] = 8000.01 >> HPI = 264
- By having K=3, the prediction for HPI is equal to the average of HPI for the top three neighbors.
- HPI = (264+139+139)/3 = 180.7

KNN for Regression

- Important thing to notice in the given data is to notice the difference in the scale of Age and Loan.

- We should ideally normalize the data in such cases.

KNN for Regression

Age	Loan	House Price Index	Distance
0.125	0.11	135	0.7652
0.375	0.21	256	0.5200
0.625	0.31	231	0.3160
0	0.01	267	0.9245
0.375	0.50	139	0.3428
0.8	0.00	150	0.6220
0.075	0.38	127	0.6669
0.5	0.22	216	0.4437
1	0.41	139	0.3650
0.7	1.00	250	0.3861
0.325	0.65	264	0.3771
0.7	**0.61**	**?**	

$$X_s = \frac{X - Min}{Max - Min}$$