Distance Functions

Minkowski Distance

$$dist(x_i, x_j) = ((|x_{i1} - x_{j1}|^n + |x_{i2} - x_{j2}|^n + \dots)^{\frac{1}{n}}$$

(x_i, x_j: vectors)

if $n = 1$ → Manhattan Distance
$n = 2$ → Euclidean Distance

Distance Functions for Binary Attributes (& Nominal)

Use a confusion matrix
i^{th} & j^{th} data points be x_i & x_j (vectors)

Data point i \ Data point j	1	0	
1	a	b	a+b
0	c	d	c+d
	a+c	b+d	a+b+c+d

a: # Number of attributes with value 1 for both data points

b: # attributes for x_i $f=1$ & x_j $f=0$ (value of f^{th} attribute in data point x_i)

c: # attributes for x_i $f=0$ x_j $f=1$

The distance becomes:

$$dist(x_i, x_j) = \frac{b+c}{a+b+c+d}$$

Nominal Attributes: Colors: Red, Blue, Yellow ← Simply assigning numbers

1 2 3

1 → Distance
2 → Distance
} this is wrong

Determine the number of values that will match in vectors.

Data points → x_i x_j

#attributes = r

values that match in x_i & x_j = q

} Distance $(x_i, x_j) = \frac{r-q}{q}$