Visual data. Assignment 5

s6sukana@uni-bonn.de - SUGAN KANAGASENTHINATHAN
s6roraoo@uni-bonn.de - ROHIL PRAKASH RAO
s6alguba@uni-bonn.de - ALBERT GUBAIDULLIN

May 2020

1 Exercise 2 (A Sports Game Using Multidimensional Scaling, 6 Points)

We will initialize values randomly: 5, 7, 10, 8 for Basketball, Football, Handball and Tennis respectively.
Absolute value error formula:

$$\frac{\partial J_{ee}}{\partial y_k} = \frac{2}{\sum_{i<k} \delta(i,k)^2} \sum_{i \neq k} (d(i,k) - \delta(i,k)) \frac{y_i - y_k}{d(i,k)}$$

1.1 Matrix 1.

Step 1. Basketball: $\frac{\partial J_{ee}}{\partial y_1} = \frac{2}{\sum_{i<k} \delta(i,k)^2} \sum_{1 \neq k} (d(1,k) - \delta(1,k)) \frac{y_1 - y_k}{d(1,k)} =$
$\frac{2}{2.41} \cdot ((2 - 0.12) \cdot ((5-7)/2)(5 - 0.3) \cdot ((5-10)/5)(3-1) \cdot ((5-8)/3)) =$
$= 0.828 \cdot (-8.58) = -7.1$

Football: $\frac{\partial J_{ee}}{\partial y_2} = \frac{2}{\sum_{i<k} \delta(i,k)^2} \sum_{2 \neq k} (d(2,k) - \delta(2,k)) \frac{y_2 - y_k}{d(2,k)} =$
$\frac{2}{2.41} \cdot ((2 - 0.12) \cdot ((7-5)/2)(3 - 0.18) \cdot ((7-10)/3)(1-0.88) \cdot ((7-8)/1)) =$
$= 0.828 \cdot (-1.06) = -0.88$

Handball: $\frac{\partial J_{ee}}{\partial y_3} = \frac{2}{\sum_{i<k} \delta(i,k)^2} \sum_{3 \neq k} (d(3,k) - \delta(3,k)) \frac{y_3 - y_k}{d(3,k)} =$
$\frac{2}{2.41} \cdot ((5 - 0.3) \cdot ((10-5)/5)(3 - 0.18) \cdot ((10-7)/3)(2-0.71) \cdot ((10-8)/2)) =$
$= 0.828 \cdot 8.81 = 7.3$

Tennis: $\frac{\partial J_{ee}}{\partial y_4} = \frac{2}{\sum_{i<k} \delta(i,k)^2} \sum_{4 \neq k} (d(4,k) - \delta(4,k)) \frac{y_4 - y_k}{d(4,k)} =$
$\frac{2}{2.41} \cdot ((3 - 1) \cdot ((8-5)/3)(1 - 0.88) \cdot ((8-7)/1)(2-0.71) \cdot ((8-10)/2)) =$
$= 0.828 \cdot 0.83 = 0.69$

Shift all point by $\alpha = 0.5$ multiplied by respective partial derivative.
Basketball = 8.55
Football = 7.44
Handball = 6.35
Tennis = 7.66

Step 2.
Basketball: $\frac{\partial J_{ee}}{\partial y_1} = 0.828 \cdot 2.79 = 2.31$

Football: $\frac{\partial J_{ee}}{\partial y_2} = 0.828 \cdot 0.58 = 0.48$

Handball: $\frac{\partial J_{ee}}{\partial y_3} = 0.828 \cdot -3.4 = -2.82$

Tennis: $\frac{\partial J_{ee}}{\partial y_4} = 0.828 \cdot -3.4 = 0.03$

New points:
Basketball = 7.4
Football = 7.2
Handball = 7.76
Tennis = 7.64

Step 3.
Basketball: $\frac{\partial J_{ee}}{\partial y_1} = 0.828 \cdot 0.77 = 0.64$

Football: $\frac{\partial J_{ee}}{\partial y_2} = 0.828 \cdot -0.02 = -0.01$

Handball: $\frac{\partial J_{ee}}{\partial y_3} = 0.828 \cdot -0.15 = -0.12$

Tennis: $\frac{\partial J_{ee}}{\partial y_4} = 0.828 \cdot -0.6 = -0.5$

New points:
Basketball = 7.08
Football = 7.21
Handball = 7.82
Tennis = 7.89

At this point errors are pretty stable to continue forward

1.2 Matrix 2.

Step 1.
Basketball: $\frac{\partial J_{ee}}{\partial y_1} = 0.828 \cdot -8.8 = -8.46$

Football: $\frac{\partial J_{ee}}{\partial y_2} = 0.828 \cdot -1.2 = -1.15$

Handball: $\frac{\partial J_{ee}}{\partial y_3} = 0.828 \cdot 8.8 = 8.46$

Tennis: $\frac{\partial J_{ee}}{\partial y_4} = 0.828 \cdot 1.2 = 1.15$

New points:
Basketball = 9.23
Football = 7.58
Handball = 5.77
Tennis = 7.42
Step 2.
Basketball: $\frac{\partial J_{ee}}{\partial y_1} = 0.828 \cdot 5.72 = 5.5$

Football: $\frac{\partial J_{ee}}{\partial y_2} = 0.828 \cdot -0.49 = -0.47$

Handball: $\frac{\partial J_{ee}}{\partial y_3} = 0.828 \cdot -5.72 = -5.5$

Tennis: $\frac{\partial J_{ee}}{\partial y_4} = 0.828 \cdot 0.49 = 0.47$

New points:
Basketball = 6.48
Football = 7.81
Handball = 8.52
Tennis = 7.19

Step 3.
Basketball: $\frac{\partial J_{ee}}{\partial y_1} = 0.828 \cdot -2.88 = -2.77$

Football: $\frac{\partial J_{ee}}{\partial y_2} = 0.828 \cdot 0.05 = 0.05$

Handball: $\frac{\partial J_{ee}}{\partial y_3} = 0.828 \cdot 2.88 = 2.77$

Tennis: $\frac{\partial J_{ee}}{\partial y_4} = 0.828 \cdot -0.05 = -0.04$

New points:
Basketball = 7.87
Football = 7.79
Handball = 7.13
Tennis = 7.21

At this point errors are pretty stable to continue forward

Since we are given normalized values of the dissimilarity matrix A and B there is no way to make an informed guess of the initial criterion used by the players, other than domain understanding. Observing dissMat A we can guess that the values were obtained using the Criteria - Ball Size (Maximum dissimilarity between a tennis ball and basketball and less between handball and football) And for dissMat B we can guess the Criteria to be - Number of Players.

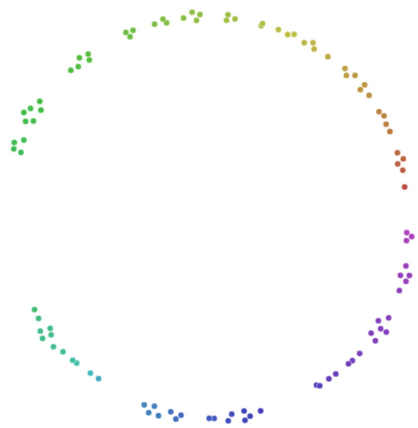

Figure 1: perplexity = 8

(Max dissimilarity between football -11 players and tennis- 2 players and less dissimilarity for basketball - 5 players and handball - 7 players)

2 Exercise 4 (Pitfalls in t-SNE, 12 Points)

1. Perplexity for N items can't be higher than N. For t-SNE this means that it needs at least 31 point to be able to use perplexity 30. If perplexity is too high, the binary search for the right bandwidth will fail and the algorithm produces wrong results.

2. Naturally, density of a square grid, is lower near the corners. With high enough perplexity, t-SNE algorithm adapts its notion of "distance" to regional density variations in the data set. So the denser regions (central parts of the grid) tend to expand and less dens regions (corners of the grid) tend to shrink.

3. In this case the perplexity of 2 seems to be too low and local variations dominate over global ones. The perplexity parameter in t-SNE sets "the effective number of neighbours" that each point is attracted to. So there will be fewer pairs that feel any attraction and the resulting embedding will tend to divide in such small clusters

4. There is no need for a high perplexity to approximate the circle. Sometimes I was able to get some approximation with perplexity equals 6 or 8 (Fig.1). Since density of the circle is equal around every point, we need a high perplexity, that considers, almost all points of circle as neighbours (Fig.2), to get a good approximation,

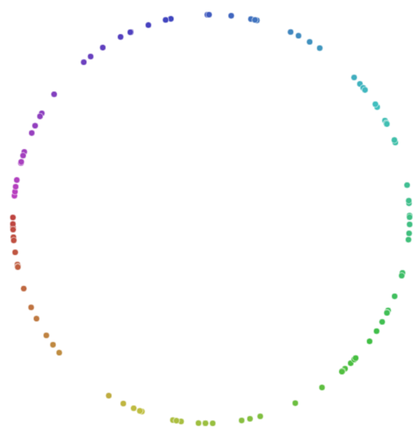

Figure 2: perplexity = 97