# **Week 5 –** **Predictive Data Analysis**

# **Exercise 01: Weighted Moving Average**

A firm has the following order history over the last 6 months.

January 120

February 95

March 100

April 75

May 100

June 50

* What would be a 3-month moving average forecast for July?

(75 + 100 + 50)/3 = 75

* What would be a 3-month weighted moving average forecast for July, using weights of 40% for the most recent month, 30% for the month preceding the most recent month, and 30% for the month preceding that one?

75\*0.3 + 100\*0.3 + 50\*0.4 = 72.5

# **Exercise 02: Exponential smoothing**

The mean price for rubber during 10 years is shown in the Table below:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Year | 2001 | 2002 | 2003 | 2004 | 2005 | 2006 | 2007 | 2008 | 2009 | 2010 |
| Price | 82 | 80 | 76 | 73 | 72 | 73 | 72 | 73 | 77 | 74 |

* Give a forecast for the price of schnaps in 2010 based on simple exponential smoothing.

Given: α = 0.2

F2010 = 0.2\*A2009 + (1-0.2)\*F2009 = 0.2\*77 + 0.2\*0.8\*73 + 0.2\*0.8\*0.8\*72 + 0.2\*0.8\*0.8\*0.8\*73 + 0.2\*0.8\*0.8\*0.8\*0.8\*72 + 0.2\*0.8\*0.8\*0.8\*0.8\*0.8\*73 + 0.2\*0.8\*0.8\*0.8\*0.8\*0.8\*0.8\*76 + 0.2\*0.8\*0.8\*0.8\*0.8\*0.8\*0.8\*0.8\*80 + 0.2\*0.8\*0.8\*0.8\*0.8\*0.8\*0.8\*0.8\*0.8\*82 = 15.4 + 11.86 + 9.216 + 7.4752 + 5.89824 + 4.784128 + 3.985 + 3.355 + 2.751 = 64.72

* Compute the Mean Absolute Deviation (MAD).

F2001 = 0

F2002 = 16.4

F2003 = 29.12

F2004 = 37.896

F2005 = 38.496

F2006 = 45.397

F2007 = 55.17

F2008 = 56.539

F2009 = 61.431

MAD = (∑|actual – forecast|)/n = 33.4017

* Do you have any viewpoint about the choice of model in this case?

With α = 0.2, the distance between forecasting value and actual value get smaller and smaller overtime (It forms a pyramid). Since the MAD is “large”, it implies that α = 0.2 is not a reliable indicator of the other value within dataset

# **Exercise 03: Items-based Recommender**

Three computers, C1, C2, and C3, have the numerical features listed below:

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature** | **Processor Speed** | **Disk Size** | **Memory Size** |
| C1 | 3.06 | 500 | 6 |
| C2 | 2.68 | 320 | 4 |
| C3 | 2.92 | 640 | 6 |

We may imagine these values as defining a vector for each computer; for instance, C1’s vector is [3.06, 500, 6]. We can compute the cosine distance between any two of the vectors, but if we do not scale the components, then the disk size will dominate and make differences in the other components essentially invisible. Let us use 1 as the scale factor for processor speed, α for the disk size, and β for the main memory size. In terms of α and β, compute the cosines of the angles between the vectors for each pair of the three computers in two following senerios:

* What are the angles between the vectors if α = β = 1?

Cos(C1, C2) = (8.2008 + 160000 + 24)/sprt(9.3636 + 250000 + 36)\*sprt(7.1824 + 102400 + 16) = 1

Cos(0) = 1, therefore angle between C1 and C2 is 0

Cos(C2, C3) = (7.8256 + 204800 + 24)/sprt(7.1824 + 102400 + 16)\*sprt(8.5264 + 409600 + 36) = 1

* Angle between C2 and C3 is 0

Cos(C1,C3) = (8.9352 + 320000 + 36)/sprt(9.3636 + 250000 + 36)\*sprt(8.5264 + 409600 + 36) = 1

* Angle between C1 and C3 is 0
* What are the angles between the vectors if α = 0.01 and β = 0.5?

Cos(C1, C2) = (8.2008 + 16+ 6)/sprt(9.3636 + 25 + 9)\*sprt(7.1824+10.24+4) = 0.99

* Angle between C1 and C2 is 8.1

Cos(C2, C3) = (7.8256+20.48+6)/sprt(7.1824+10.24+4)\*sprt(8.5264+40.96+9) = 34.3056/4.628\*7.647 = 0.97

* Angle between C2 and C3 is 14

Cos(C1, C3) = (8.9352+32+9)/sprt(9.3636+25+9)/sprt(8.5264+40.96+9) = 49.93/6.585\*7.648 = 0.99

* Angle between C1 and C3 is 8.1
* Do you have any viewpoint about the choice of α, β in this case?

If α = β, angle between each vectors is always equal 0 (means that the vector is the same direction and thus perfect similarity). The smaller the value of α and β are, the larger the angle between the two vectors