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

# **Exercise 01: Weighted Moving Everage**

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?

+ 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?

+ July = 0.4\*50 + 0.3\*100 + 0.3\*75 = 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.
* Compute the Mean Absolute Deviation (MAD).
* Do you have any viewpoint about the choice of model in this case?

1. = 0.2

F2001 = 0

F2002 = 0.2\*82 + 0.8\*F2001 = 16.4

F2003 = 0.2\*80 + 0.8\*F2002 = 29.12

F2004 = 0.2\*76 + 0.8\*F2003 = 38.496

F2005 = 0.2\*73 + 0.8\*F2004 = 45.397

F2006 = 0.2\*72 + 0.8\*F2005 = 50.71

F2007 = 0.2\*73 + 0.8\*F2006 = 55.71

F2008 = 0.2\*72 + 0.8\*F2007 = 58.539

F2009 = 0.2\*73 + 0.8\*F2008 = 61.431

………………

F2010 = 0.2\*77 + 0.8\*F2009 = 64.56

1. Compute the Mean Absolute Deviation (MAD)

Mean of the set = = 75.2

Sum of |data value – mean| = 6.8 + 4.8 + 0.8 + 2.2 + 3.2 + 2.2 + 3.2 + 2.2 + 1.8 + 1.2 = 28.4

MAD = = = 2.84

# **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?

For

+ cos (A, B) = = 1

cos 0 = 1, the angle is 0

+ cos (B, C) = = 1

cos 0 = 1, the angle is 0

+ cos (A, C) = = 1

cos 0 = 1, the angle is 0

* What are the angles between the vectors if α = 0.01 and β = 0.5?
* + cos (A, B) = = 0.99
* cos 0 = 1, the angle is 0
* + cos (B, C) = = 0.96
* cos 0 = 1, the angle is 0
* + cos (A, C) = = 0.99
* cos 0 = 1, the angle is 0
* Do you have any viewpoint about the choice of α, β in this case?

We should choose small for the disk size not dominating the others since the disk size values are very big comparing to process speed and memory size, and small is not recommended since the memory size values are not so big.