Assignment 3: Clarifications and Test Dataset

Explanation for Mahalanobis Norm:

For calculating the covariance matrix, we need {xi-mu} to be mutually independent for all i=1, 2, 3 … p (where p is the dimension of each vector).

In the additional data (4th argument of the subroutine), pass X = [x1; x2; x3; ... xp], such that {x1 - mu, x2 - mu, ..., xp - mu} forms a set of mutually independent vectors, where mu = (x1+x2+....+xp)/p

Mahalanobis Norm can also be written as:

$$\text{Mahalanobis Norm} = ||x||_A = \text{sqrt}(x.A.(x)^T)$$

$$A = \text{inv}(\text{Cov}\, X) = \text{inv}\left((1/(p-1)) * \sum_{i=1\ldots p} (xi - mu)^T(xi - mu) \right)$$

$$= \text{inv}\left((1/(p-1)) * ((X')^T.(X')) \right)$$

where; T: Transpose, inv: Inverse, Cov: Covariance, (.): Dot Product

X' = [x1 – mu; x2 – mu; ....; xp – mu]

Some Examples to verify code:

1. Euclidean Norm

x	y	Out
[4, 12, 3, 5]	[9, 8, 2, 3]	(6.782329983125268, 0.12849622184722817)

2. Frobenius or Hilbert Schmidt Norm

x	y	Out
[4, 12, 3, 5]	[9, 8, 2, 3]	(2.0, 0.3333333333333333)

3. Diagonal Norm

x	y	diagonal	Out
[4, 12, 3, 5]	[9, 8, 2, 3]	[7, 1, 10, -1]	(14.035668847618199, 0.06650851452866428)
[1, 2, 4]	[-6, 0, 9]	[1, 1, 0, 1]	"The vector dimension isn't compatible"
[1,-2]	[-11, 4]	[-11, 4]	"Dissimilarity turns out to be the square root of a negative number"

4. Mahalanobis Norm

x	y	X	Out
[4, 12, 3, 5]	[9, 8, 2, 3]	[[4,18,21,37], [14,7,2,11], [3,5,2,13], [11,34,47,26], [71, 2, 1, 15]]	(3.5102967568357557, 0.22171490123890655)

[61, 53, 8]	[7, -2, 10]	[[-13,2,1,6], [17,5,-8,2], [44, 23,73,-11]]	“The shape of distribution does not conform with the data points”
[15, 41]	[68, -29]	[[-1,1], [0,0]]	“The distribution results in a non-invertible covariance matrix”

5. Lebesgue or Minkowski Norm

x	y	alpha	Out
[4, 12, 3, 5]	[9, 8, 2, 3]	6	(5.2006744337772055, 0.16127277938552234)
[2, 13, 7, 11, 5]	[-8, 31, 1, 6, 52]	0	"Similarity/ Dissimilarity measures not defined”

6. Cosine

x	y	Out
[4, 12, 3, 5]	[9, 8, 2, 3]	(0.14429494027266943, 0.8739005695172557)
[3, -2, 4]	[4, -2, -4]	(inf, 0)

7. Overlap

x	y	Out
[4, 12, 3, 5]	[9, 8, 2, 3]	(0.03267973856209139, 0.9683544303797469)
[0, 0, 0]	[4, -2, -4]	“Measures not defined; similarity has 0/0 form”

8. Dice

x	y	Out
[4, 12, 3, 5]	[9, 8, 2, 3]	(0.15032679738562105, 0.8693181818181818)

9. Jaccard

x	y	Out
[4, 12, 3, 5]	[9, 8, 2, 3]	(0.30065359477124187, 0.7688442211055276)