



	Improved method for correcting sample Mahalanobis distance without estimating population eigenvalues or eigenvectors of covariance matrix
	International Journal of Data Science and Analytics, Vol. 10, Issue 2, pp 121-134 (2020) 電子版 : https://link.springer.com/article/10.1007/s41060-019-00201-4
	Yasuyuki Kobayashi
	<p>The recognition performance of the sample Mahalanobis distance (SMD) deteriorates as the number of learning samples decreases. Therefore, it is important to correct the SMD for a population Mahalanobis distance (PMD) such that it becomes equivalent to the case of infinite learning samples. In order to reduce the computation time and cost for this main purpose, this paper presents a correction method that does not require the estimation of the population eigenvalues or eigenvectors of the covariance matrix. In short, this method only requires the sample eigenvalues of the covariance matrix, number of learning samples, and dimensionality to correct the SMD for the PMD. This method involves the summation of the SMD's principal components (each of which is divided by its expectation obtained using the delta method), Lawley's bias estimation, and the variances of the sample eigenvectors. A numerical experiment demonstrates that this method works well for various cases of learning sample number, dimensionality, population eigenvalues sequence, and non-centrality. The application of this method also shows improved performance of estimating a Gaussian mixture model using the expectation-maximization algorithm.</p>