














3 Decorrelating your data and dimension reduction

0%

Dimension reduction summarizes a dataset using its common occurring patterns. In this chapter, you'll learn about the most fundamental of dimension reduction techniques, "Principal Component Analysis" ("PCA"). PCA is often used before supervised learning to improve model performance and generalization. It can also be useful for unsupervised learning. For example, you'll employ a variant of PCA will allow you to cluster Wikipedia articles by their content!

 Visualizing the PCA transformation	50 xp
 Correlated data in nature	100 xp
 Decorrelating the grain measurements with PCA	100 xp
 Principal components	50 xp
 Intrinsic dimension	50 xp
 The first principal component	100 xp
 Variance of the PCA features	100 xp
 Intrinsic dimension of the fish data	50 xp
 Dimension reduction with PCA	50 xp
 Dimension reduction of the fish measurements	100 xp
 A tf-idf word-frequency array	100 xp
 Clustering Wikipedia part I	100 xp
 Clustering Wikipedia part II	100 xp

HIDE CHAPTER DETAILS