Mahalanobis Distance

Uri Shaham

October 11, 2022

1 Introduction

Proposition 1.1. *Let $x_1, \ldots, x_n$ be datapoints in $\mathbb{R}^m$, and let C be the $m \times m$ data covariance matrix $C = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(x_i - \bar{x})^T$. Then C is semi positive-definite. If, in addition, $n \geq m$ and $x_1, \ldots, x_n$ span $\mathbb{R}^m$, then C is strictly positive-definite.*

Proof. Pick $y \in \mathbb{R}^m$, and let $z_i = x_i - \bar{x}$. Then $y^T C y = \frac{1}{n}\sum_{i=1}^{n}\left(z_i^T y\right)^T\left(z_i^T y\right) \geq 0$, so C is positive semi-definite. Now assume that $x_1, \ldots, x_n$ span $\mathbb{R}^m$ (then so do $z_1, \ldots z_n$). If y is nonzero, then $y^T C y = 0$ implies that $z_i^T y = 0$ for all i, which is a contradiction, as for any linear combination $\sum_i a_i z_i$, $y^T \sum_i a_i z_i = 0 \neq y^T y$. □

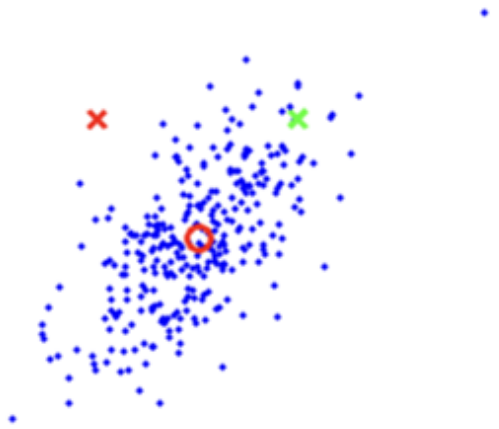

Figure 1: Which of the points marked in x is closer to the cluster centroid? in what sense?

The main takeout from the above example is that distance should be data driven, and take the distribution of the data into account. Mahalanobis distance takes into account the covariance of the data, by multiplying the Euclidean distance with the inverse covariance. Specifically, for two points $x_i, x_j \in \mathbb{R}^m$, the distance is defined as

$$d_M(x_i, x_j) = \sqrt{(x_i - x_j)^T C^{-1}(x_i - x_j)}.$$

Since C is typically positive definite (for $n \geq m$), it can be inverted, so the distance is well-defined. To understand the Mahalanobis distance, consider the eigendecomposition $C = V\Lambda V^T$. Let $W = \Lambda^{-\frac{1}{2}} V^T$ be the (PCA) whitening matrix. Then $TC^{-1} = W^T W$. So the Mahalanobis distance is

$$d_M(x_i, x_j) = \|W(x_i - x_j)\|$$

, i.e., the Euclidean distance between the whitened points Wx_i and Wx_j. Thus Mahalanobis distance is in fact the standard Euclidean distance on whitened data.

2 Local Mahalanobis

Real world data often lies on low dimensional manifold (e.g., a spiral). Many times, we are interested in finding neighboring points (for example, in applications of kNN. In such cases, it is benefitial to take into account local covariance matrices, rather than the global covariance matrix. For example, data on a cross (isotropic covariance). This yields the local Mahalanobis distance, where for each point we compute neighbors using its local metric, defined using the local covariance matrix. This can be used to design an iterated kNN algorithm as follows. In each iterations we find the k nearest neighbors using the current local metric (starting with the Euclidean metric at the first iteration). Then we compute the local covariance, and use the local metric to find new neighbors, The process is repeated until convergence (i.e., neighbors don't change).

Homework

1. Show that Mahalanobis distance with uncorrelated features is equal to Euclidean distance after standardization.
2. Generate data and perform nearest iterated nearest neighbor search using local Mahalanobis distance. Present the results in an insightful way.

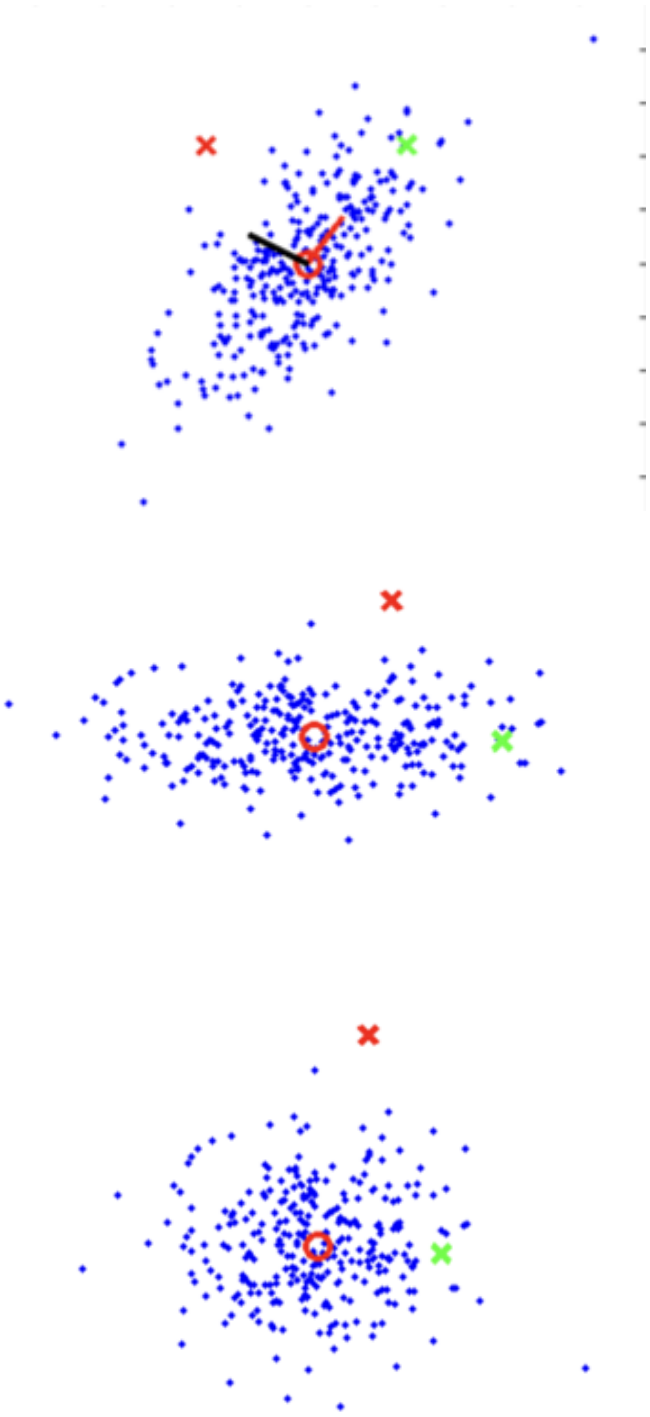

Figure 2: Mahalanobis distance is Euclidean distance after whitening.; Top: principal directions. Center: projection onto the principal directions. Bottom - whitening (Identity covariance)