Decorrelation and Whitening

Yun Shen

November 18, 2015

Let $\{x_i \in \mathbb{R}^m, i = 1, \dots, n\}$ be n m-dimensional input data. We assume each x_i is a column vector. Suppose here $n > m$. Let

$$X := [x_1, x_2, \dots, x_n]$$

be the matrix such that its ith column is x_i. See Figure 1 for one example.

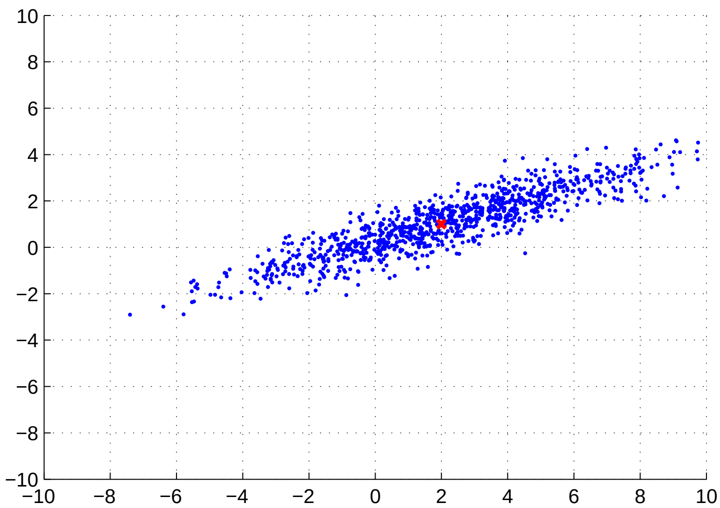

Figure 1: Original data points by blue dots. $n = 1000$, $m = 2$. Red cross represents the center $x^{(c)}$ of all data points, cf. Equation (1).

Centering We can calculate the center vector $x^{(c)} \in \mathbb{R}^m$ of all data points as follows

$$x^{(c)} := \frac{1}{n} \sum_{i=1}^{n} x_i \tag{1}$$

We can then center the data by

$$\bar{x}_i := x_i - x^{(c)}$$

which satisfy that

$$\frac{1}{n} \sum_{i=1}^{n} \bar{x}_i = 0.$$

Figure 2 plot the centered data points. Let $\bar{X} := [\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n]$ be the corresponding matrix.

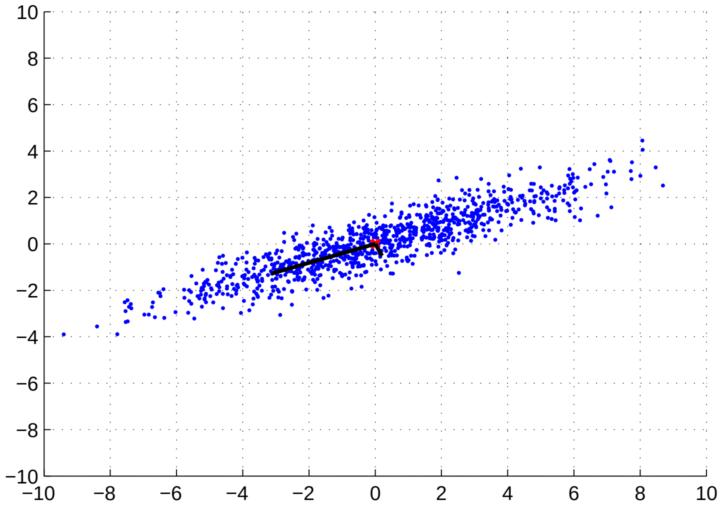

Figure 2: Data points $\{\bar{x}_i\}$ after centering. The center now is $(0,0)$ marked by the red cross.

Decorrelation After centering, we can calculate the covariance matrix

$$C := \frac{1}{n}\sum_{i=1}^{n} \bar{x}_i \bar{x}_i^\top = \frac{1}{n}\bar{X}\bar{X}^\top.$$

For more detailed introduction see wikipedia page of the *principal component analysis*. In essence, the (k,l) element of the matrix C, C_{kl}, calculates the covariance between the kth and lth elements of all data points. As we have seen in the last tutorial, C is a symmetric, positive semi-definite matrix. However, the non-diagonal elements of C are usually non-zero, i.e., the kth and lth $(k \neq l)$ elements are *correlated*.

Hence, the task of decorrelation is to find a matrix $V \in \mathbb{R}^{m\times m}$ such that Λ is diagonal, where

$$\Lambda = \frac{1}{n}\sum_{i=1}^{n} \tilde{x}_i \tilde{x}_i^\top, \text{ and } \tilde{x}_i := V^\top \bar{x}_i. \tag{2}$$

How can we find such matrix V? We can simply perform an eigenvalue decomposition of C. Because C is positive semi-definite, it has the following representation

$$C = V\Lambda V^\top, \tag{3}$$

where $V = [v_1, v_2, \ldots, v_m]$ is composed of *orthonormal* eigenvectors $v_i (i = 1, \ldots, m)$ of C and $\Sigma = diag(\lambda_1, \lambda_2, \ldots, \lambda_m)$ is the diagonal matrix of eigenvalues $\lambda_i (i = 1, \ldots, m)$. It is easy to verify that Equation (3) implies Equation (2). So V is exactly the matrix we want!

Figure 3 plots the distribution of $\{\tilde{x}_i\}$ after decorrelation. Comparing with the data points before decorrelation shown in Figure 2, decorrelation simply "rotates" the data points such that the two directions (plotted by black lines) coincide with the horizontal and vertical axis, of which consist the *trivial* orthonormal basis of the Euclidean space.

Whitening We can see in Figure 3 that the width of the data distribution in the horizontal direction is much smaller than the width in the vertical direction. In whitening,

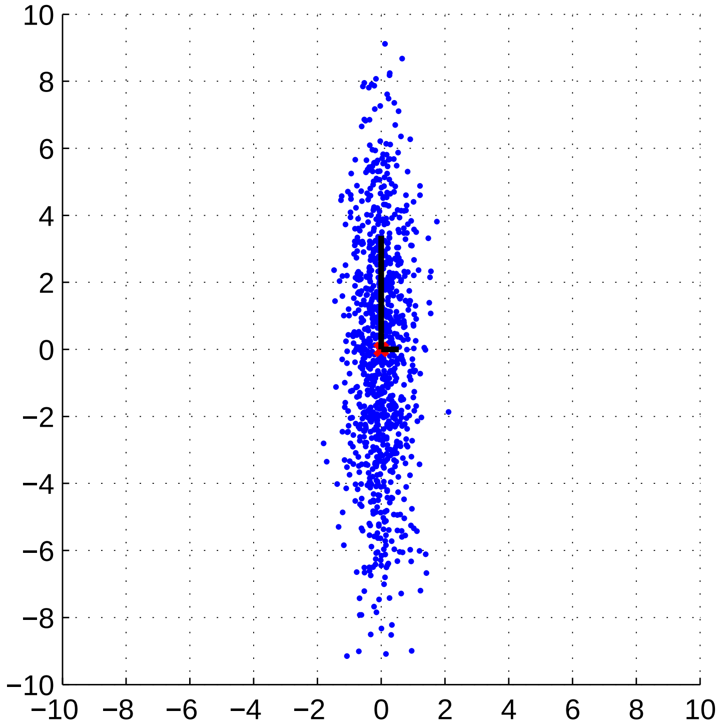

Figure 3: Data points $\{\tilde{x}_i\}$ after decorrelation.

we rescale the data such that the width in each direction is identical. We do the following: let

$$\hat{x}_i := \Lambda^{-1/2}\tilde{x}_i \tag{4}$$

where $\Lambda^{-1/2}$ is the diagonal matrix of $(\lambda_1^{-1/2}, \lambda_2^{-1/2}, \ldots, \lambda_m^{-1/2})$ with λ_i being the eigenvalues of matrix C (see also Equation (3)). It is easy to verify that

$$\frac{1}{n}\sum_{i=1}^{n} \hat{x}_i\hat{x}_i^\top = I$$

the identity matrix. Figure 4 plots the distribution of $\{\hat{x}_i\}$ after whitening. Clearly the data distribution has an identical width in each direction.

Summary

1. Centering, $\bar{x}_i = x_i - \frac{1}{n}\sum_i x_i$
2. Decorrelation, $\tilde{x}_i = V^\top \bar{x}_i$, where V is the eigenvector-matrix of $C = \frac{1}{n}\sum_i \bar{x}_i\bar{x}_i^\top$
3. Whitening, $\hat{x}_i = \Lambda^{-1/2}V^\top \bar{x}_i$, where Λ is the eigenvalue-matrix of C

Questions Q1: Suppose we have a training data set $\{x_i\}_{i=1}^n$ together with labels $\{y_i\}$. We first decorrelate (or whiten) data and obtain $\{\tilde{x}_i\}$ (or $\{\hat{x}_i\}$) and use them for training. After training, we obtain a prediction function f using the decorrelated data as input. The procedure can be illustrated as follows.

$$\text{(input) } x_i \xrightarrow{\text{decorrelation}} \tilde{x}_i \xrightarrow{f} \hat{y}_i \text{ (prediction)}$$

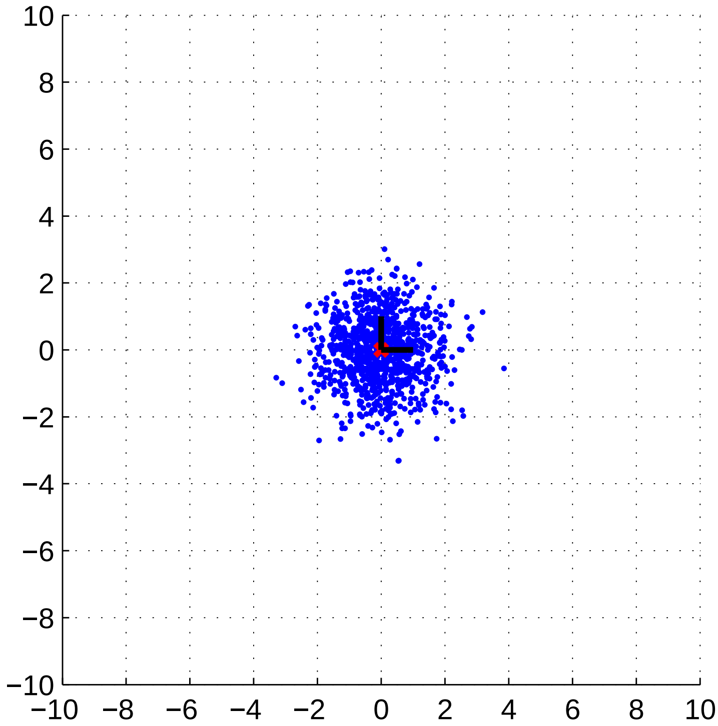

Figure 4: Data points $\{\hat{x}_i\}$ after whitening.

How would you decorrelate (or whiten) a new input x^*, which is not within the training data set?

Q2: How would you deal with the case where some of the eigenvalues of C are zero?