Topic: Decorrelating and then Whitening data

Extra notes for MAS622J/1.126J by Rosalind W. Picard

1. Let $\mathbf{x}$ be a vector of zero-mean data. Form its covariance matrix,

$$\mathbf{\Sigma} = E(\mathbf{x}\mathbf{x}^T)$$

   If the data points in $\mathbf{x}$ are correlated, then their covariance, $\mathbf{\Sigma}$, will NOT be a diagonal matrix.

2. In order to decorrelate the data, we need to transform it so that the transformed data will have a diagonal covariance matrix. This transform can be found by solving the eigenvalue problem. We find the eigenvectors and associated eigenvalues of $\mathbf{\Sigma}$ by solving

$$\mathbf{\Sigma}\mathbf{\Phi} = \mathbf{\Phi}\mathbf{\Lambda}$$

   $\mathbf{\Lambda}$ is a diagonal matrix having the eigenvalues as its diagonal elements.

   The matrix $\mathbf{\Phi}$ thus diagonalizes the covariance matrix of $\mathbf{x}$. The columns of $\mathbf{\Phi}$ are the eigenvectors of the covariance matrix.

   We can also write the diagonalized covariance as:

$$\mathbf{\Phi}^T\mathbf{\Sigma}\mathbf{\Phi} = \mathbf{\Lambda} \tag{1}$$

   If we wish to apply this diagonalizing transform to a single vector of data we just form:

$$\mathbf{y} = \mathbf{\Phi}^T\mathbf{x} \tag{2}$$

   Thus, the data $\mathbf{y}$ has been decorrelated: its covariance, $E[\mathbf{y}\mathbf{y}^T]$ is now a diagonal matrix, $\mathbf{\Lambda}$.

3. The diagonal elements (eigenvalues) in $\mathbf{\Lambda}$ may be the same or different. If we make them all the same, then this is called *whitening* the data. Since each eigenvalue determines the length of its associated eigenvector, the covariance will correspond to an ellipse when the data is not whitened, and to a sphere (having all dimensions the same length, or uniform) when the data is whitened. Whitening is easy:

$$\mathbf{\Lambda}^{-1/2}\mathbf{\Lambda}\mathbf{\Lambda}^{-1/2} = \mathbf{I}$$

   Equivalently, substituting in (1), we write:

$$\mathbf{\Lambda}^{-1/2}\mathbf{\Phi}^T\mathbf{\Sigma}\mathbf{\Phi}\mathbf{\Lambda}^{-1/2} = \mathbf{I}$$

Thus, to apply this whitening transform to $\mathbf{y}$ we simply multiply it by this scale factor, obtaining the whitened data $\mathbf{w}$:

$$\mathbf{w} = \mathbf{\Lambda}^{-1/2}\mathbf{y} = \mathbf{\Lambda}^{-1/2}\mathbf{\Phi}^T\mathbf{x} \tag{3}$$

Now the covariance of $\mathbf{w}$ is not only diagonal, but also uniform (white), since the covariance of $\mathbf{w}$, $\mathrm{E}(\mathbf{w}\mathbf{w}^T) = \mathbf{I}$:

$$E(\mathbf{\Lambda}^{-1/2}\mathbf{\Phi}^T\mathbf{x}\mathbf{x}^T\mathbf{\Phi}\mathbf{\Lambda}^{-1/2}) = \mathbf{I}$$

In DHS, the diagonalizing transform applied to $\mathbf{x}$ is denoted $\mathbf{A}$ and the whitening transform is represented by $\mathbf{A_w}$. These map onto the notation above as follows:

$$\mathbf{y} = A^T\mathbf{x}, \ \ A = \mathbf{\Phi}$$

$$\mathbf{w} = A_w^T\mathbf{x}, \ \ A_w = \mathbf{\Phi}\mathbf{\Lambda}^{-1/2}$$