Lecture 5

1 Eigenvalues and Eigenvectors of matrices

Now that we know how to project our data set onto our new axes, the next problem we have is knowing how to find our direction vectors for the new axis in the first place. To do this, we need to learn about an important mathematical concept, which is the *eigenvectors* and the corresponding *eigenvalues* for a (square and symmetric) matrix.

Definition 1.1. Let $\mathbf{A} \in \mathbb{R}^{d \times d}$ be a $d \times d$ square symmetric matrix. A vector $\mathbf{v} \in \mathbb{R}^d$ is called an *eigenvector* of $\mathbf{A}$ if

$$\mathbf{A}\mathbf{v} = \lambda \mathbf{v}.$$

The scalar $\lambda \in \mathbb{R}$ is called *eigenvalue* of $\mathbf{A}$.

This concept of vectors that only 'stretch' or 'contract' under the transformation by a matrix is important because it can be shown that[1]:

a) A $d \times d$ matrix will have exactly d eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_d$, with $\lambda_1, \lambda_2, \ldots, \lambda_d$ as the corresponding eigenvalues[2].

b) For *any* square matrix $\mathbf{A}$, we can decompose it into a very special form, known as *matrix signalisation*:

$$\mathbf{A} = \mathbf{P}\mathbf{D}\mathbf{P}^{-1}.$$

$\mathbf{P}$ is a matrix formed by the eigenvectors of $\mathbf{A}$, and $\mathbf{D}$ is a diagonal matrix where the non-zero entries contain the eigenvalues of $\mathbf{A}$.

2 Decomposition of the Co-variance Matrix

With these facts we can now find a relationship between the co-variance matrix $\mathbf{C}$ in the original axes $(x_1, x_2, \ldots, x_n)$ and the co-variance matrix $\tilde{\mathbf{C}}$ in our new axis $(\tilde{x}_1, \tilde{x}_2, \ldots, \tilde{x}_n)$, where we we would like $\tilde{\mathbf{C}}$ to be a diagonal matrix with the only (potentially) non-zero entries on the diagonal: $Var[\tilde{X}_1], Var[\tilde{X}_2], \ldots, Var[\tilde{X}_n]$. Using fact b), we can show that

$$\mathbf{C} = \mathbf{V}\tilde{\mathbf{C}}\mathbf{V}^T,$$

---

[1]see http://mathworld.wolfram.com/MatrixDiagonalization.html for details
[2]Strictly speaking this is true only if all eigenvalues are non-zero and different.

where $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n]$, each $\mathbf{v}_i$ being the direction vector for axis $\tilde{x}_i$ and an eigenvector of $\mathbf{C}$. This also means that $Var[\tilde{X}_i]$ is the corresponding eigenvalue of $\mathbf{C}$ i.e. $\mathbf{C}\mathbf{v}_i = Var[\tilde{X}_i]\mathbf{v}_i$.

Remember that the direction vectors have the following properties:

- $||\mathbf{v}_i|| = 1$ for every i,

- $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ whenever $i \neq j$.

To further demonstrate the arguments above, suppose that we are in two-dimensional space. We want to show that $\mathbf{C}\mathbf{a} = \mathbf{V}\tilde{\mathbf{C}}\mathbf{V}^T\mathbf{a}$ for every $\mathbf{a} \in \mathbb{R}^2$. It suffices to show that $\mathbf{C}\mathbf{v}_i = \mathbf{V}\tilde{\mathbf{C}}\mathbf{V}^T\mathbf{v}_i$ for $i = 1, 2$, with $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2]$ and $\tilde{\mathbf{C}} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$.

We know that $\mathbf{C}\mathbf{v}_i = \lambda_i\mathbf{v}_i$. Now,

$$
\begin{aligned}
\mathbf{V}\tilde{\mathbf{C}}\mathbf{V}^T\mathbf{v}_1 &= \mathbf{V}\tilde{\mathbf{C}} \begin{bmatrix} \mathbf{v}_1^T\mathbf{v}_1 \\ \mathbf{v}_2^T\mathbf{v}_1 \end{bmatrix} \\
&= \mathbf{V}\tilde{\mathbf{C}} \begin{bmatrix} 1 \\ 0 \end{bmatrix} \\
&= \mathbf{V} \begin{bmatrix} \lambda_1 \\ 0 \end{bmatrix} \\
&= \lambda_1\mathbf{v}_1
\end{aligned}
$$

Using similar reasoning, $\mathbf{C}\mathbf{v}_2 = \mathbf{V}\tilde{\mathbf{C}}\mathbf{V}^T\mathbf{v}_2$.

3 Principal Component Analysis (PCA)

We finally have the tools we need in order to perform PCA on a given data set $\mathcal{D} = \{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^N\}$ in d-dimensional space:

1) Using the data, calculate the co-variance matrix $\mathbf{C}$.

2) Calculate the eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_d$ of $\mathbf{C}$ (vectors for our new axes) and their corresponding eigenvalues $\lambda_1, \ldots, \lambda_d$ (the variability along each new axis), such that $||\mathbf{v}_i|| = 1$ for every i and $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d \geq 0$.[3]

3) Pick the first k of those eigenvectors such that a proportion of variability is 'sufficiently' preserved. We can find the fraction of data variability preserved in the first k dominant axes, ρ, using the following expression:

$$
\rho = \frac{\lambda_1 + \lambda_2 + \ldots + \lambda_k}{\lambda_1 + \lambda_2 + \ldots + \lambda_d}
$$

Suppose that we want to preserve a proportion ρ' of variability. Then choose k eigenvectors such that $\rho \geq \rho'$.

---

[3]This ordering can be achieved in packages like Octave or MATLAB which saves us the task of doing it manually.

4) Project our data onto the new axes by using

$$(\mathbf{V}^{(k)})^T \mathbf{x}^i = \tilde{\mathbf{x}}^i,$$

where $\mathbf{V}^{(k)} = [\mathbf{v}_1, \ldots, \mathbf{v}_k]$.