ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Unsupervised and reinforcement learning in neural networks
Week 1

Lecturer: Marc-Oliver Gewaltig
Assistants: Dane Corneil
Berat Denizdurduran

Solutions – Linear Algebra Revision

September 17, 2014

Exercise 1: Eigenvalue Problems

The number λ is called the eigenvalue of A *if and only if*

$$det(A - \lambda I) = 0 \tag{1}$$

This is the characteristic equation. So, each eigenvalue, λ, has a corresponding eigenvector x

$$(A - \lambda I)x = 0 \quad or \quad Ax = \lambda x \tag{2}$$

Let's start to find the eigenvalues of A, given that

$$A = \begin{bmatrix} 2 & 0 \\ 0 & -1 \end{bmatrix}$$

In this example, we need to shift the A by λI according to Eq.1

$$A - \lambda I = \begin{bmatrix} 2-\lambda & 0 \\ 0 & -1-\lambda \end{bmatrix}$$

So, the determinant of $A - \lambda I$ is the characteristic polynomial

$$(2-\lambda)(-1-\lambda) = 0$$

The root of the characteristic polynomial yields the eigenvalues

$$\lambda_1 = 2 \quad and \quad \lambda_2 = -1$$

Now find the eigenvectors separately for $\lambda_1 = 2$ and $\lambda_2 = -1$ using Eq. 2

$$\lambda_1 = 2: \qquad (A - \lambda_1 I)x = \begin{bmatrix} 0 & 0 \\ 0 & -3 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

The first eigenvector is now

$$x_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$\lambda_2 = -1 : \qquad (A - \lambda_2 I)x = \begin{bmatrix} 3 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

The second eigenvector is now

$$x_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

Vectors are orthogonal *if and only if* their dot product is zero: $v \cdot w = 0$ *or* $v^T w = 0$

$$x_1 \cdot x_2 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \cdot \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 0$$

So, the eigenvectors of A are orthogonal to each other and since any eigenvector may be multiplied by a scalar, it's not unique.
Solutions for the rest of the matrices;

$$B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \lambda_{1,2} = 1 \quad x_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad x_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, x_1 \cdot x_2 = 0$$

$$C = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \lambda_1 = 1 \quad \lambda_2 = -1 \quad x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad x_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}, x_1 \cdot x_2 = 0$$

$$D = 1/2 \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} \quad \lambda_1 = 1 \quad \lambda_2 = 0 \quad x_1 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} \quad x_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, x_1 \cdot x_2 = 0$$

$$E = \begin{bmatrix} 2 & -1 \\ 0 & 1 \end{bmatrix} \quad \lambda_1 = 2 \quad \lambda_2 = 1 \quad x_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix} \quad x_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, x_1 \cdot x_2 \neq 0$$

Exercise 2: From Data to the Covariance Matrix

First we note that the mean of the data vanishes. Then we see due to the point symmetry we only have to compute half of the points. Finally the axial symmetry makes the $\text{var}(\mathbf{x}_1) = \text{var}(\mathbf{x}_2)$. So we have

$$\text{var}(\mathbf{x}_1) = \text{cov}(\mathbf{x}_1, \mathbf{x}_1) = \frac{2}{8}[1 + 1 + 4 + 4] = \frac{20}{8}$$

and

$$\text{cov}(\mathbf{x}_1, \mathbf{x}_2) = \frac{2}{8}[1 + 2 + 2 + 4] = \frac{18}{8}$$

so in total

$$C = \frac{1}{8}\begin{pmatrix} 20 & 18 \\ 18 & 20 \end{pmatrix}$$

Exercise 3: Properties of Covariance Matrices

1. By defining $\mathbf{y}^\mu = (\mathbf{x}^\mu - \langle \mathbf{x} \rangle)$ we see

$$\mathbf{C} = \frac{1}{M}\sum_{\mu=1}^{M} \mathbf{y}^\mu \mathbf{y}^{\mu T}$$

and $(\mathbf{y}^\mu \mathbf{y}^{\mu T})^T = (\mathbf{y}^{\mu TT} \mathbf{y}^{\mu T}) = (\mathbf{y}^\mu \mathbf{y}^{\mu T})$ and therefore $\mathbf{C}$ is a sum of symmetric matrices and hence symmetric itself. Positive semindefinite we show along similar lines

$$\mathbf{v}^T \mathbf{y}^\mu \mathbf{y}^{\mu T} \mathbf{v} = (\mathbf{v}^T \mathbf{y}^\mu)(\mathbf{v}^T \mathbf{y}^\mu)^T = \left(\mathbf{v}^T \mathbf{y}^\mu\right)^2 \geq 0$$

and consequently $\mathbf{C}$, the sum of positive semidefinite quantities, is also positive semidefinite.

2. From the previous result, we note that

$$\mathbf{n}^T \mathbf{C} \mathbf{n} = \frac{1}{M} \sum_{\mu=1}^{M} \left(\mathbf{n}^T \mathbf{y}^\mu\right)^2$$

Since $\mathbf{n}$ is normalized, the term inside the brackets corresponds to the scalar projection of the data point $\mathbf{y}^\mu$ on to the vector $\mathbf{n}$. By squaring and averaging these projections, we get the variance of the N–dimensional data set when it is projected on to the 1–dimensional line defined by $\mathbf{n}$.

Exercise 4: Projectors

1. $\mathbf{A}$ projects $\mathbf{x}$ onto $\mathbf{n}$.

2. If vector $\mathbf{x}$ lies in the subspace that $\mathbf{P}$ projects to, then $\mathbf{P}$ does not have an effect on $\mathbf{x}$ and $\mathbf{Px} = \mathbf{x}$. Therefore with $\mathbf{x} = \mathbf{Py}$ for any vector $\mathbf{y}$ we have $\mathbf{P}^2 = \mathbf{P}$.

3. We have to show $\mathbf{A}^2 = \mathbf{A}$

$$\mathbf{A}^2 = \mathbf{AA} = \mathbf{nn}^T \mathbf{nn}^T = \mathbf{n}(\mathbf{n}^T \mathbf{n})\mathbf{n}^T = \mathbf{nn}^T = \mathbf{A}$$

4. For instance

$$\begin{aligned}
\mathbf{P} &= (\mathbf{aa}^T + \mathbf{bb}^T) \\
\mathbf{P}x &= (\mathbf{aa}^T + \mathbf{bb}^T)x \\
\mathbf{P}x &= \mathbf{a}(\mathbf{a}^T x) + \mathbf{b}(\mathbf{b}^T x)
\end{aligned}$$

Since the bracketed terms are scalars, we have some combination of the vectors $\mathbf{a}$ and $\mathbf{b}$. Note that this is not necessarily a projector since, in general, $\mathbf{P} \neq \mathbf{P}^2$. If, however, $\mathbf{a}$ and $\mathbf{b}$ are normalized and orthogonal, we have

$$\begin{aligned}
\mathbf{P}^2 &= (\mathbf{aa}^T + \mathbf{bb}^T)(\mathbf{aa}^T + \mathbf{bb}^T) \\
&= \mathbf{aa}^T \mathbf{aa}^T + \mathbf{aa}^T \mathbf{bb}^T + \mathbf{bb}^T \mathbf{aa}^T + \mathbf{bb}^T \mathbf{bb}^T \\
&= \mathbf{aIa}^T + \mathbf{a0b}^T + \mathbf{b0a}^T + \mathbf{bIb}^T \\
&= \mathbf{aa}^T + \mathbf{bb}^T = \mathbf{P}
\end{aligned}$$

5. B and D