Machine Learning I: Foundations
Exercise Sheet 6

Prof. Marius Kloft TA:Billy Joe Franks

10.06.2020
Deadline: 09.06.2020

1) (MANDATORY) 10 Points

Let $M \in \mathbb{R}^{n \times n}$ be a symmetric square matrix. Prove that the following statements are equivalent.

- M is positive semi-definite, i.e. $\forall \mathbf{z} \in \mathbb{R}^n : \mathbf{z}^T M \mathbf{z} \geq 0$.
- All eigenvalues of M are non-negative.

As we are proving equivalence we will show from the first statement follows the second ($\Rightarrow$) and vice versa ($\Leftarrow$).

$\Rightarrow$ Let M be positive semi-definite. Consider arbitrary eigenvalues λ and corresponding eigenvector $\mathbf{v}_\lambda$, i.e. $M\mathbf{v}_\lambda = \lambda \mathbf{v}_\lambda$.

$$0 \leq \mathbf{v}_\lambda^T M \mathbf{v}_\lambda = \mathbf{v}_\lambda^T \lambda \mathbf{v}_\lambda = \lambda \mathbf{v}_\lambda^T \mathbf{v}_\lambda = \lambda \left\| \mathbf{v}_\lambda \right\|^2$$

As such $\lambda \geq 0$.

$\Leftarrow$ Let $\lambda_1, \dots, \lambda_n$ be the non-negative eigenvalues of M. Since M is a symmetric matrix we can find eigenvectors which form an orthogonal basis, this is a well known property of symmetric matrices. This follows from a symmetric matrix being normal and using the spectral theorem. It is not necessary for the course to understand the above, however it can be a good practice to try and follow along. Anyway now that we have an orthogonal basis of eigenvectors consider any $\mathbf{z} \in \mathbb{R}^n$, we can represent $\mathbf{z}$ in the new basis, i.e. $\mathbf{z} = \sum_i a_i \mathbf{v}_{\lambda_i}$. Now consider $\mathbf{z}^T M \mathbf{z}$

$$\mathbf{z}^T M \mathbf{z} = \left(\sum_i a_i \mathbf{v}_{\lambda_i}^T \right) M \left(\sum_i a_i \mathbf{v}_{\lambda_i} \right) = \left(\sum_i a_i \mathbf{v}_{\lambda_i}^T \right) \left(\sum_i a_i \lambda_i \mathbf{v}_{\lambda_i} \right).$$

Now consider that the basis was orthogonal, as such if $i \neq j$ then $\mathbf{v}_{\lambda_i}^T \mathbf{v}_{\lambda_j} = 0$ from which we can follow

$$\left(\sum_i a_i \mathbf{v}_{\lambda_i}^T \right) \left(\sum_i a_i \lambda_i \mathbf{v}_{\lambda_i} \right) = \sum_i a_i^2 \lambda_i \mathbf{v}_{\lambda_i}^T \mathbf{v}_{\lambda_i} = \sum_i a_i^2 \lambda_i \left\| \mathbf{v}_{\lambda_i} \right\|^2.$$

$a_i^2 \geq 0, \lambda_i \geq 0, \|\mathbf{v}_{\lambda_i}\|^2 \geq 0$, so the whole thing is non-negative.

$$\Rightarrow \mathbf{z}^T M \mathbf{z} \geq 0$$

2) A square matrix M is called diagonalizable if and only if there exists an invertible matrix P such that $P^{-1}MP = D$ with D diagonal, i.e. $D = diag(\mathbf{v})$ for some $\mathbf{v}$.

a) Prove that M is diagonalizable if M has n distinct eigenvalues.

Since M has n distinct eigenvalues, each of these must have a unique eigenvector. If some of these vectors were linearly dependent, i.e. $\sum_{i \in I} a_i \mathbf{v}_{\lambda_i} = \mathbf{v}\lambda_j$, then

$$\begin{aligned} M\mathbf{v}_{\lambda_j} &= \lambda_j \mathbf{v}_{\lambda_j} \\ M \sum_{i \in I} a_i \mathbf{v}_{\lambda_i} &= \lambda_j \sum_{i \in I} a_i \mathbf{v}_{\lambda_i} \\ \sum_{i \in I} a_i \lambda_i \mathbf{v}_{\lambda_i} &= \lambda_j \sum_{i \in I} a_i \mathbf{v}_{\lambda_i} \end{aligned}$$

However this implies $\forall i \in I : \lambda_i = \lambda_j$ which contradicts the assumption. As such the vectors must be linearly independent. Now consider the definition of eigenvalues

$$M\mathbf{v}_{\lambda_i} = \lambda_i \mathbf{v}_{\lambda_i} = \mathbf{v}_{\lambda_i} \lambda_i.$$

We can construct a matrix P where each column of P is one of the eigenvectors, then $MP = PD$, where D is a diagonal matrix containing all the eigenvalues. As the eigenvectors are linearly independent the rank of P is n and P is invertible, so $P^{-1}MP = D$ and we have proven the statement.

b) The converse of the above is not true. Give an example for this.

Consider the identity I. All of its eigenvalues are 1 and thus equal. However, it can be diagonalized using itself, i.e. $P = I$, $D = I$ then $I^{-1}II = I$.

3) Are the following matrices diagonalizable? If yes, determine P and D as above. If no, give a reason why not.

a)

$$\begin{bmatrix} 0 & 1 \\ -8 & 4 \end{bmatrix}$$

Diagonalizable, however imaginary.

$$D = \begin{pmatrix} 2+2i & 0 \\ 0 & 2-2i \end{pmatrix}, P = \begin{pmatrix} \frac{i}{2} & \frac{1}{8} - i\frac{1}{8} \\ -\frac{i}{2} & \frac{1}{8} + i\frac{1}{8} \end{pmatrix}.$$

b)

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

$$D = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{pmatrix}, P = \begin{pmatrix} -\frac{1}{3} & \frac{2}{3} & -\frac{1}{3} \\ -\frac{1}{3} & -\frac{1}{3} & \frac{2}{3} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \end{pmatrix}.$$

c)

$$\begin{bmatrix} 5 & 4 & 2 & 1 \\ 0 & 1 & -1 & -1 \\ -1 & -1 & 3 & 0 \\ 1 & 1 & -1 & 2 \end{bmatrix}$$

Not diagonalizable. It has the 3 eigenvalues, $1, 2, 4$, where 4 has multiplicity 2. This is where we have to search to figure out wether it is diagonalizable. In fact the eigenvalue 4 only has 1 corresponding eigenvector, thus we will not be able to construct the invertible matrix P, note the invertability is the problem, we would have to use 4 twice in D with the same corresponding eigenvector, this would lead to P not having full rank and thus not being invertible. This is typicaly expressed as: the eigenvalue 4 does not have the same geometric and algebraic multiplicity, thus the matrix is not diagonalizable.

d)

$$\begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}$$

$$D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{pmatrix}, P = \begin{pmatrix} -1 & 2 & 2 \\ -1 & 3 & 2 \\ 3 & -6 & -5 \end{pmatrix}.$$

4) Solve programming task 6.