Cognitive Algorithms - Exercise Sheet 2

Solutions * Exercises

Department of Machine Learning - TU Berlin

Task 3 - PSD: Pretty Sweet, Dude! [*]

We care in particular about PSD matrices because all covariance matrices are PSD. One interesting property about PSD matrices is that all of their eigenvalues are non-negative. In particular it holds that all real, symmetric matrices A with non-negative eigenvalues are PSD, and vice versa. Prove this statement in both directions!

Hint 1: For the direction

$$\text{PSD} \Rightarrow \text{Eigenvalues are positive}$$

start with $A\mathbf{v} = \lambda\mathbf{v}$, for some eigenvector $\mathbf{v}$ of A.

Hint 2: For the direction

$$\text{Eigenvalues are positive} \Rightarrow \text{PSD}$$

consider that for all real symmetric matrices A, they can be decomposed using an eigenvalue decomposition, such that $A = Q\Lambda Q^T$, where Q is an orthogonal matrix and Λ contains the eigenvalues of A

Solutions

1. To Prove: PSD $\Rightarrow$ Eigenvalues ≥ 0. Assume: M is PSD (and symmetric by definition). Consider some eigenvector-eigenvalue pair $\mathbf{v}, \lambda$ of M:

$$\begin{aligned}
A\mathbf{v} = \lambda\mathbf{v} &\implies \mathbf{v}^T A\mathbf{v} = \mathbf{v}^T \lambda\mathbf{v} \\
&\implies \mathbf{v}^T A\mathbf{v} = \lambda\mathbf{v}^T\mathbf{v} \\
&\implies \mathbf{v}^T A\mathbf{v} = \lambda\mathbf{v}^2 \\
&\implies \frac{\mathbf{v}^T A\mathbf{v}}{\mathbf{v}^2} = \lambda \\
&\overset{A \text{ is } PSD \implies \mathbf{v}^T A\mathbf{v} \geq 0 \text{ and } \forall x \in .x^2 \geq 0}{\implies} \lambda \geq 0 \quad \square
\end{aligned}$$

2. To Prove: Real symmetric and Eigenvalues $\geq 0 \implies$ PSD.
Assume: all Eigenvalues of $M \geq 0$ and M symmetric, real. Consider some

eigenvector-eigenvalue pair $\mathbf{v}, \lambda$ of M:

$$A \text{ real, symmetric } \implies \exists Q, \Lambda, \text{ where } Q \text{ orthogonal and } \Lambda \text{ diagonal with all Eigenvalues of A on diagonal}$$

$$\text{such that } Q\Lambda Q^T = A \overset{P=Q^T}{\implies} P^T \Lambda P = A$$
$$\implies \mathbf{v}^T P^T \Lambda P \mathbf{v} = \mathbf{v}^T A \mathbf{v} \text{ for some arbitrary, non-zero } \mathbf{v} \in^d$$
$$\implies (P\mathbf{v})^T \Lambda P \mathbf{v} = \mathbf{v}^T A \mathbf{v}$$
$$\overset{\text{Define } y=P\mathbf{v}}{\implies} y^T \Lambda y = \mathbf{v}^T A \mathbf{v}$$
$$\overset{\Lambda \text{ is diagonal}}{\implies} \sum_{i=1}^{d} y_i \Lambda_{i,i} y_i = \mathbf{v}^T A \mathbf{v}$$
$$\overset{\lambda_i \text{ is eigenvalue of A}}{\implies} \sum_{i=1}^{d} y_i \lambda_i y_i = \mathbf{v}^T A \mathbf{v} \implies \sum_{i=1}^{d} y_i^2 \lambda_i = \mathbf{v}^T A \mathbf{v}$$
$$\overset{\text{Assumption: A has non-neg eigenvalues and } \forall x \in . x^2 \geq 0}{\implies} 0 \leq \mathbf{v}^T A \mathbf{v}$$
$$\implies A \text{ is PSD} \quad \square$$

Task 5 - Whitening [*]

Whitening is a linear transformation of a data set. Its purpose is to decorrelate data and set the variance to one, featurewise. Thus, when whitening is applied, the covariance matrix is the identiy matrix afterwards. For many algorithms, whitening is a useful preprocessing step.

Suppose we have n empirical data points $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \mathbb{R}^d$ with zero mean. The covariance matrix is $\Sigma_X = \frac{1}{n} X X^T$. It can be rewritten as $\Sigma_X = U \Lambda U^T$. This is the eigendecomposition into an orthogonal matrix U with eigenvectors in the columns and a diagonal matrix Λ with the eigenvalues on the diagonal. Remember that for orthogonal matrices U the inverse is the transposed and thus $U^T U = U U^T = I$.

In this exercise we use the whitening operation as the mapping

$$\mathbf{z}_k := P^T \mathbf{x}_k = U \Lambda^{-\frac{1}{2}} U^T \mathbf{x}_k = \Sigma_X^{-\frac{1}{2}} \mathbf{x}_k$$

To get a geometrical intuition take a look at figures 1 and 2. Note that there are different ways to define whitening. For example one could also skip the last rotation by the matrix U visualized in figure 1.

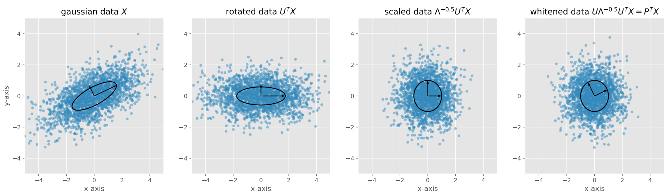

Figure 1: The whitening process: the different matrices are applied one after another.

1. Show that P is a symmetric matrix.
2. Show that
$$\Sigma_X^{\frac{1}{2}} = U \Lambda^{\frac{1}{2}} U^T$$
is a valid square root of a positive definite matrix.
3. Show that P is a valid inverse of the square root of a positive definite matrix.

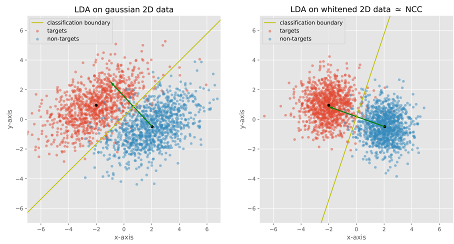

Figure 2: LDA applied to 2D gaussian toy data with and without whitening

4. Show that the covariance of the whitened data P^TX is the identity matrix. Keep in mind that we deal with data with zero mean.
5. Proof that classification with LDA is equivalent to classification with NCC of the whitened data.

Solutions

1.

$$\begin{aligned} P^T &= (U\Lambda^{-\frac{1}{2}}U^T)^T \\ &= U^{T^T}\Lambda^{(-\frac{1}{2})^T}U^T \\ &= U\Lambda^{-\frac{1}{2}}U^T \end{aligned}$$

2.

$$\begin{aligned} \Sigma_X^{\frac{1}{2}}\Sigma_X^{\frac{1}{2}} &= U\Lambda^{\frac{1}{2}}U^TU\Lambda^{\frac{1}{2}}U^T \\ &= U\Lambda^{\frac{1}{2}}\Lambda^{\frac{1}{2}}U^T \\ &= U\Lambda U^T \\ &= \Sigma_X \end{aligned}$$

3.

$$\begin{aligned} PP &= \Sigma_X^{-\frac{1}{2}}\Sigma_X^{-\frac{1}{2}} \\ &= U\Lambda^{-\frac{1}{2}}U^TU\Lambda^{-\frac{1}{2}}U^T \\ &= U\Lambda^{-\frac{1}{2}}\Lambda^{-\frac{1}{2}}U^T \\ &= U\Lambda^{-1}U^T \\ &= \Sigma_X^{-1} \end{aligned}$$

4.

$$\begin{aligned} \Sigma_{P^TX} &= \frac{1}{n}(P^TX)(P^TX)^T \\ &= \frac{1}{n}P^TXX^TP \\ &= P^T\Sigma_XP \\ &= P^T(U\Lambda U^T)P \\ &= P^T(U\Lambda U^T)P \\ &= P(U\Lambda U^T)P \\ &= U\Lambda^{-\frac{1}{2}}U^T(U\Lambda U^T)U\Lambda^{-\frac{1}{2}}U^T \\ &= U\Lambda^{-\frac{1}{2}}\Lambda\Lambda^{-\frac{1}{2}}U^T \\ &= UU^T \\ &= I \end{aligned}$$

5. Since we have zero mean, no bias is needed for classification.

$$\begin{aligned}
\mathbf{w}_{nCC}^T \mathbf{z}_k &= (P^T \bar{\mathbf{x}}_+ - P^T \bar{\mathbf{x}}_-)^T P^T \mathbf{x}_k \\
&= P^T (\bar{\mathbf{x}}_+ - \bar{\mathbf{x}}_-)^T P^T \mathbf{x}_k \\
&= (\bar{\mathbf{x}}_+ - \bar{\mathbf{x}}_-) P P^T \mathbf{x}_k \\
&= (\bar{\mathbf{x}}_+ - \bar{\mathbf{x}}_-) P P \mathbf{x}_k \\
&= (\bar{\mathbf{x}}_+ - \bar{\mathbf{x}}_-) \Sigma_X^{-1} \mathbf{x}_k \\
&= {\Sigma_X^{-1}}^T (\bar{\mathbf{x}}_+ - \bar{\mathbf{x}}_-)^T \mathbf{x}_k \\
&= \Sigma_X^{-1} (\bar{\mathbf{x}}_+ - \bar{\mathbf{x}}_-)^T \mathbf{x}_k \\
&= \mathbf{w}_{LDA}^T \mathbf{x}_k
\end{aligned}$$

@write“endstandalonepreambles@out