Mathematics of Learning – Worksheet 3

Basics. [Singular Value Decomposition].
Calculate the SVD of the matrix

$$A = \begin{pmatrix} 2 & -2 & -2 & 0 \\ -1 & -1 & 3 & 4 \\ 2 & -2 & 2 & -2 \end{pmatrix}.$$

The SVD of a matrix $A \in \mathbb{R}^{m\times n}$ consists of matrices $U \in \mathbb{R}^{m\times m}, V \in \mathbb{R}^{n\times n}$ which are orthogonal, i.e., UU^T, VV^T is equal to the corresponding unit matrix, and a matrix $\Sigma \in \mathbb{R}^{m\times n}$ which has only positive, descending entries on the diagonal (it looks like

$$\begin{pmatrix} * & 0 & 0 & 0 \\ 0 & * & 0 & 0 \\ 0 & 0 & * & 0 \end{pmatrix}$$

in our case). $U \cdot \Sigma \cdot V^T = A$ should hold, in case you calculated correctly. Hint: No nice numbers this time.

Solution. SVDs are calculated computationally by sophisticated numerical algorithms, which also work efficient in the case the matrix A is large or ill-conditioned. A possibility to calculate the SVD manually on a sheet of paper is the following: U can be chosen to be the matrix of (normalized) eigenvectors of AA^T. The vectors have to be sorted according to the size of the corresponding eigenvalues. If v is a normalized eigenvector of AA^T, it is not clear beforehand if to take v or $-v$ since both would be normalized eigenvectors. Analogously, V can be chosen to be the matrix of normalized eigenvectors of A^TA. Hence, U and V are equal to (for example, every column might be multiplied by -1)

$$U = \begin{pmatrix} 0.3559 & 0.3110 & -0.8813 \\ -0.9050 & 0.3499 & -0.2420 \\ 0.2331 & 0.8837 & 0.4060 \end{pmatrix}, \quad V = \begin{pmatrix} -0.3811 & 0.5003 & 0.2429 & -0.7385 \\ 0.0499 & -0.6720 & -0.4088 & -0.6155 \\ 0.5417 & 0.5385 & -0.6336 & -0.1231 \\ 0.7476 & -0.0902 & 0.6102 & -0.2462 \end{pmatrix}$$

Σ is a "diagonal" matrix with the squareroots of the eigenvalues of A^TA (or AA^T, respectively; it is the same).

$$\Sigma = \begin{pmatrix} 5.4658 & 0 & 0 & 0 \\ 0 & 4.0762 & 0 & 0 \\ 0 & 0 & 2.9171 & 0 \end{pmatrix}$$

Calculating the product $U\Sigma V^T$, we see that it is not equal to A. We have to set the signs right, using the identity (u_k denoting the k-th column of U, and v_k denoting the k-th column of V, and σ_k denoting the k-th singular value):

$$u_k = \sigma_k^{-1} A v_k$$

Hence, we derive (and technically, this makes it useless that we calculated a candidate for U beforehand)

$$U = \begin{pmatrix} -0.3559 & 0.3110 & 0.8813 \\ 0.9050 & 0.3499 & 0.2420 \\ -0.2331 & 0.8837 & -0.4060 \end{pmatrix}$$

Multiplying $U\Sigma V^T$ we get A.

Exercise 1 [Reading assignment: Association rules].
Read chapter 14.2 of the *Hastie* book, regarding association rules. Discuss the contents with one (or more) fellow student for at least half an hour.

Exercise 2 [Reading assignment: Self organizing maps (SOM)].
Read chapter 14.5 of the *Hastie* book, regarding self organizing maps. Discuss the contents with one (or more) fellow student for at least half an hour.

Exercise 3 [Prerequisites for PCA].
Given a set of data vectors $x_1, ..., x_N \in \mathbb{R}^p$ and a matrix $V_q \in \mathbb{R}^{p\times q}$, $q < p$, with q orthogonal unit vectors as columns. Prove, that

$$\tilde{\mu} = \bar{x}, \quad \tilde{\lambda}_i = V_q^T(x_i - \bar{x})$$

is a minimizer (over μ and λ_i)

$$\sum_{i=1}^{N} ||x_i - \mu - V_q\lambda_i||^2,$$

where $||\cdot||$ denotes the euclidean norm. Furthermore, show that the minimizer $\bar{x}$ for μ is not unique and find the set of minimizers for μ.

Solution. To calculate the minimizer, we calculate the first derivative of the function, i.e.,

$$\frac{\partial}{\partial \mu_k} \sum_{i=1}^{N}\sum_{j=1}^{p}(x_{i,j} - \mu_j - (V_q\lambda_i)_j)^2 = \sum_{i=1}^{N} 2(x_{i,k} - \mu_k - (V_q\lambda_i)_k),$$

evaluating to zero in case we use the proposed values for $\tilde{\mu}$ and $\tilde{\lambda}_i$:

$$\sum_{i=1}^{N} 2(x_i - \bar{x} - V_qV_q^T(x_i - \bar{x})) = 2(\sum_{i=1}^{N} x_i - N\bar{x} - V_qV_q^T(\sum_i x_i - N\bar{x})) = 0,$$

since $\sum_{i=1}^{N} x_i - N\bar{x}$ is the zero vector. The minimizer is not unique, since we can choose any vector for μ such that $\sum_{i=1}^{N} x_i - N\mu$ is an eigenvector corresponding to eigenvalue 1 of the matrix $V_qV_q^T$.

On the other hand, if we calculate the partial derivative for $\lambda_{k,m}$, we get

$$\frac{\partial}{\partial \lambda_{k,m}} \sum_{i=1}^{N}\sum_{j=1}^{p}(x_{i,j} - \mu_j - (V_q\lambda_i)_j)^2 = \sum_{j=1}^{p} 2(x_{k,j} - \mu_j - (V_q\lambda_k)_j)\cdot(V_q)_{j,m} =$$

$$\langle x_k - \mu - V_q\lambda_k, (V_q)_{\cdot,m}\rangle = \langle x_k - \mu - V_q\lambda_k, (V_q)e_m\rangle =$$
$$\langle V_q^T(x_k - \bar{x}) - V_q^T V_q V_q^T(x_k - \bar{x}), e_m\rangle = 0.$$

The hessian matrices are nonnegative diagonal matrices with similar arguments than for k-means-clustering, additionally utilizing the orthogonality of V_q.