Machine Learning Exercise Sheet 10

Dimensionality Reduction & Matrix Factorization

In-class Exercises

Problem 1: In this exercise, we use proof by induction to show that the linear projection onto an M-dimensional subspace that maximizes the variance of the projected data is defined by the M eigenvectors of the data covariance matrix S, given by

$$S = \frac{1}{N}\sum_{n=1}^{N}(\boldsymbol{x}_n - \bar{\boldsymbol{x}})(\boldsymbol{x}_n - \bar{\boldsymbol{x}})^T \qquad \bar{\boldsymbol{x}} = \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}_n$$

corresponding to the M largest eigenvalues. In Section 12.1 in Bishop this result was proven for the case of $M = 1$. Now suppose the result holds for some general value of M and show that it consequently holds for dimensionality $M + 1$. To do this, first set the derivative of the variance of the projected data with respect to a vector $\boldsymbol{u}_{M+1}$ defining the new direction in data space equal to zero. This should be done subject to the constraints that $\boldsymbol{u}_{M+1}$ be orthogonal to the existing vectors $\boldsymbol{u}_1, \ldots \boldsymbol{u}_M$, and also that it be normalized to unit length. Use Lagrange multipliers to enforce these constraints. Then make use of the orthonormality properties of the vectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_M$ to show that the new vector $\boldsymbol{u}_{M+1}$ is an eigenvector of S. Finally, show that the variance is maximized if the eigenvector is chosen to be the one corresponding to eigenvector λ_{M+1} where the eigenvalues have been ordered in decreasing value.

Suppose that the result holds for projection spaces of dimensionality M. The $M + 1$ dimensional principal subspace will be defined by the M principal eigenvectors $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_M$ together with an additional direction vector $\boldsymbol{u}_{M+1}$ whose value we wish to determine. We must constrain $\boldsymbol{u}_{M+1}$ such that it cannot be linearly related to $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_M$ (otherwise it will lie in the M-dimensional projection space instead of defining an $M + 1$ independent direction). This can easily be achieved by requiring that $\boldsymbol{u}_{M+1}$ be orthogonal to $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_M$, and these constraints can be enforced using Lagrange multipliers $\eta_1, ..., \eta_M$.

Following the argument given in section 12.1.1 for $\boldsymbol{u}_1$ we see that the variance in the direction $\boldsymbol{u}_{M+1}$ is given by $\boldsymbol{u}_{M+1}^T \boldsymbol{S} \boldsymbol{u}_{M+1}$. We now maximize this using a Lagrange multiplier λ_{M+1} to enforce the normalization constraint $\boldsymbol{u}_{M+1}^T \boldsymbol{u}_{M+1} = 1$. Thus we seek a maximum of the function:

$$\boldsymbol{u}_{M+1}^T \boldsymbol{S} \boldsymbol{u}_{M+1} + \lambda_{M+1}(1 - \boldsymbol{u}_{M+1}^T \boldsymbol{u}_{M+1}) + \sum_{i=1}^{M} \eta_i \boldsymbol{u}_{M+1}^T \boldsymbol{u}_i$$

with respect to $\boldsymbol{u}_{M+1}$. The stationary points occur when

$$0 = 2\boldsymbol{S}\boldsymbol{u}_{M+1} - 2\lambda_{M+1}\boldsymbol{u}_{M+1} + \sum_{i=1}^{M} \eta_i \boldsymbol{u}_i$$

Left multiplying with $\boldsymbol{u}_j^T$, and using the orthogonality constraints, we see that $\eta_j = 0$ for $j = 1, \ldots, M$. We therefore obtain

$$\boldsymbol{S}\boldsymbol{u}_{M+1} = \lambda_{M+1}\boldsymbol{u}_{M+1}$$

and so $\boldsymbol{u}_{M+1}$ must be an eigenvector of $\boldsymbol{S}$ with eigenvalue λ_{M+1}. The variance in the direction $\boldsymbol{u}_{M+1}$ is given by $\boldsymbol{u}_{M+1}^T\boldsymbol{S}\boldsymbol{u}_{M+1} = \lambda_{M+1}$ and so is maximized by choosing $\boldsymbol{u}_{M+1}$ to be the eigenvector having the largest eigenvalue amongst those not previously selected. Thus the result holds also for projection spaces of dimensionality $M+1$, which completes the inductive step. Since we have already shown this result explicitly for M = 1 if follows that the result must hold for any $M \ll D$.

Problem 2: Proof that minimizing the error is equivalent to maximizing the variance.

See Bishop chapter 12.1.2.