14.15 (Multi-Layer Support Vector Machines •) *By first extracting nonlinear principal components according to (14.16), and then training a Support Vector Machine, we can construct Support Vector type machines with additional layers. Discuss the architecture, and the different ways of training the different layers.*

14.16 (Mechanical Analogy ◦◦◦) *Try to generalize the mechanical PCA algorithm described in [443], which interprets PCA as an iterative spring energy minimization procedure, to a feature space setting. Try to come up with mechanically inspired ways of taking into account negative data in PCA (cf.* oriented *PCA, [140]).*

14.17 (Kernel PCA and Locally Linear Embedding ••) *Suppose we approximately represent each point of the dataset as a linear combination of its n nearest neighbors. Let $(W_n)_{ij}$, where $i, j \in [m]$, be the weight of point x_j in the expansion of x_i minimizing the squared representation error.*

1. *Prove that $k_n(x_i, x_j) := \left((\mathbf{1} - W_n)^\top(\mathbf{1} - W_n)\right)_{ij}$ is a positive definite kernel on the domain $\mathcal{X} = \{x_1, \ldots, x_m\}$.*

2. *Let λ be the largest eigenvalue of $(\mathbf{1} - W_n)^\top(\mathbf{1} - W_n)$. Prove that the* LLE kernel $k_n^{\mathrm{LLE}}(x_i, x_j) := \left((\lambda - 1)\mathbf{1} + W_n^\top + W_n - W_n^\top W_n\right)_{ij}$ *is positive definite on $\{x_1, \ldots, x_m\}$.*

3. *Prove that kernel PCA using the LLE kernel provides the LLE embedding coefficients [445] for a d-dimensional embedding as the coefficient eigenvectors $\boldsymbol{\alpha}^2, \ldots, \boldsymbol{\alpha}^{d+1}$. Note that if the eigenvectors are normalized in $\mathcal{H}$, then dimension i will be scaled by $\lambda_i^{-1/2}$, $i = 1, \ldots, d$.*

4. *Discuss the variant of LLE obtained using the centered Gram matrix*

$$(\mathbf{1} - 1_m)\left((\lambda - 1)\mathbf{1} + W_n^\top + W_n - W_n^\top W_n\right)(\mathbf{1} - 1_m) \tag{14.47}$$

(cf. (14.17)). Show that in this case, the LLE embedding is provided by $\alpha^1, \ldots, \alpha^d$.

5. *Interpret the LLE kernel as a similarity measure based on the similarity of the coefficients required to represent two patterns in terms of n neighboring patterns.*

14.18 (Optimal Approximation Property of PCA •) *Discuss whether the solutions of KFA satisfy the optimal approximation property of Proposition 14.1.*

14.19 (Scale Invariance ••) *Show that the problems of Kernel PCA and Sparse Kernel Feature Analysis are scale invariant; meaning that the solutions for $\Omega[f] \le c$ and $\Omega[f] \le c'$ for $c, c' > 0$ are identical up to a scaling factor.*

Show that this also applies for a rescaling of the data in Feature Space. What happens if we rescale in input space? Analyze specific kernels such as $k(x, x') = \langle x, x' \rangle^d$ and $k(x, x') = \exp(-\frac{\|x - x'\|^2}{2\sigma^2})$.

14.20 (Contrast Functions for Projection Pursuit •••) *Compute for $q(\xi) = \xi^4$ the expectations under a normal distribution of unit variance. What happens if you use a different distribution with the same variance?*