Exercise 1. Contour lines in Kernel PCA

To get a better sense of the effect of using a kernel in Kernel PCA and other methods, we will plot the contour lines that are created using two classical kernels for a few casebased examples. The contour lines (or level curves) denote all the points that have equal value on the projection onto the eigenvectors in feature space. The projections are given by:

$$< v^i, \phi(x) >= \sum_{j=1}^{M} \alpha_j^i k(x^j, x), \tag{1}$$

where $x^j \in \mathbb{R}^2,\ j = 1...M$ are M training datapoints, $v^i,\ i = 1...M$, are the M eigenvectors with dual eigenvectors α^i. Write the Gram matrix, compute the dual eigenvectors and draw the isolines of the equation 1 for the projection on each eigenvector:

1) Gaussian Kernel: $k(x, x') = e^{-\frac{||x-x'||^2}{2\sigma^2}}, \sigma \in \mathbb{R}$
   i. With $M = 1$ datapoint
   ii. With two datapoints
   iii. With three equidistant datapoints and $k(x^i, x^j) = 0.5$

   Discuss in each case the effect of the choice of the kernel width σ.

2) Polynomial Kernel: $k(x, x') =< x, x' >^p$
   i. $M = 1$ (assume data is not centered!)
   ii. With $p = 2$ and $M = 2$ (consider the case where data are centered and when not)
   iii. With $p = 2$ and $M = 3$ (and the 3 points are equidistant)
   iv. What can you say about more general cases with arbitrary M and p?

Exercise 2. Kernel PCA – data centering

Kernel PCA like PCA assumes that the data $X = \{x^i\}_{i=1}^M$ are centered, i.e. $< X >= 0$. Assume that the data are not centered prior to computing the Gram matrix .

i. How can you now modify to make sure that the projections in the feature space have zero mean? Recall that each element of the Gram matrix computes the inner product across two data points, i.e. $K_{ij} =< \phi(x^i), \phi(x^j) >$.

ii. True or False? If K is a positive definite kernel matrix, then all of its entries are positives.

Exercise 3. Matrix Decompositions

A: $A = \begin{bmatrix} -1 & 2 \\ -7 & 8 \end{bmatrix}$.

i) Compute the eigenvectors and eigenvalues of A.
ii) Using matlab, compute the spectral decomposition of A.

B: Recall that the eigenvectors of a matrix A form an orthonormal basis. If A is a square 2 by 2 matrix, its spectral decomposition can be written as $A = V\Lambda V^{-1}$, where V is a matrix, whose columns are the 2 eigenvectors of A and Λ contains the eigenvalues. V^{-1} can be thought of as a rotation. If you take a vector x and apply V^{-1}, $y = V^{-1}x$, y is an image of x through a rotation V^{-1}.

i) Using the eigenvectors found in A and for $x = \begin{bmatrix}1\\1\end{bmatrix}$, draw in the plane the different images x',x'',x''' of x when computing $x' = V^{-1}x$, $x'' = \Lambda x'$, $x''' = Vx''$.

ii) What can you say about the different transformations?

Exercise 4. Probabilistic PCA – derivation (optional)

Probabilistic PCA assumes that the N-dim. observed dataset $X = \{x^i\}_{i=1}^M$ was generated by an underlying process modeled as:

$$X = W^T z + \mu + N(0, \sigma_\varepsilon^2),$$

where z are a set of $q-$dimensional ($q < N$) latent variables,μ - an offset and σ_ε^2 - the variance of the zero mean Gaussian noise. Show that the log likelihood estimate of the projection matrix W is given by:

$$\log(L(B, \mu)) = -\frac{M}{2}(N \ln(2\pi) + \ln|B| + tr(B^{-1}C)),$$

where C is the covariance matrix of X and $B = WW^T + \sigma_\varepsilon^2 I$.