Problem 7.1

A kernel $K(x, y)$ is called a Mercer kernel if it satisfies the Mercer conditions: 1) $K(x, y)$ is symmetric, 2) continuous, and 3) positive semi-definite. Prove the following results

(a) If $K_1$ and $K_2$ are Mercer kernels, then their product, $K = K_1K_2$ is also a Mercer kernel.
(b) If $K_1$ and $K_2$ are Mercer kernels, then their linear combination, $K = aK_1 + bK_2$, $a \geq 0$ and $b \geq 0$ is also a Mercer kernel.
(c) $K(x, y) = (x^Ty + c)^d$ is a Mercer kernel for any $c \geq 0$ and for $d = 1, 2, ...$ [Hint: use the results from (a) and (b)].
(d) Let $x, y \in \{1, 2, ..., 100\}$. Show that $K(x, y) = \min(x, y)$ is a positive-definite discrete kernel. That is, $\sum_{x,y} f(x)f(y)K(x, y) \geq 0$ for any nonzero sequence $f(x)$, $x = 1, 2, ..., 100$.

Matlab Exercises

Problem 7.2

Again, for the diabetes dataset, create a training set and a test set as you did in HWs 4, 5, 6. You are expected to turn in original code for each part, but you may reuse your own code from previous assignments.

(a) Use PCA to reduce the dimensionality of the training set to two features; train a linear classifier on the new, lower dimensional training set. Test your classifier on the test set (use the transformations learned on the training set to project the test set onto the lower dimensional space).
(b) Learn the weights of an autoencoder (implemented with a restricted Boltzmann machine) with two nodes in the hidden layer (excluding bias) on the training set; train a linear classifier on the lower-dimensional space represented by the hidden nodes. Test your classifier on the test set.
(c) Implement kernel PCA with a Gaussian kernel to reduce the dimensionality of the training set to two features. As before, learn a linear classifier on the training set and test the classifier on the test set.
(d) Comment on your results, comparing them to each other and also to a linear classifier trained on the original feature space.