Homework III

Kernels and Statistical Learning Theory

Due: 27th of March, before class

Problem 1 (20%)

Consider a similarity measure k: X -> {0,1}. Show that, if

- $k(x,x) = 1$ for all x

then k is a kernel if and only if, for all x, x', x''

- $k(x,x') = 1 \Leftrightarrow k(x',x) = 1$
- $k(x,x') = k(x,x'') = 1 \Leftrightarrow k(x,x'') = 1$

Problem 2 (20%)

Note that the kernel from Problem 1 corresponds to an equivalence relation $R \subseteq X \times X$ so that $k(x,x') = 1$ if and only if $(x,x') \in R$. Imagine you could pick any kernel of this type. What kernel would be ideal for a given training and test set? Explain your intuition.

Problem 3 (40%)

Compare a linear SVM to an SVM with a kernel on the dataset on the "liver" dataset available on the course home page. The task is to predict liver disorder based on blood test and alcohol consumption. The task has six (normalized) attributes and the data is split into a training and a test set. Compare the results of a soft-margin SVM with polynomial kernel

$$K(a,b)= (a*b+0)^d$$

to those of a linear soft-margin SVM. To specify the polynomial kernel in svm_learn, use the option –t 1, pick the degree with –d <value>, and select the constant term with –r 0. Try degrees 1 to 4. Make sure you try an appropriate range of values for C using the option –c <value>.

Discuss and try to explain the results with respect to all interesting differences and trends you observe. Consider at least the training error, the test error, and the value of the optimal C.

Problem 4 (20%)

Consider a learning problem in which there is only one real-valued feature $x \in \Re$, and C=H is the set of intervals over the reals, $H=\{(a < x < b) \mid a,b \in \Re\}$. What is the probability that a hypothesis consistent with n examples of this target concept will have error at least ε? Solve this using the VC dimension.

Also, find a second way to solve this, based on first principles and ignoring the VC dimension.