

CMPS 242 Third Homework, Fall 2006

Just exercises, not to be turned in.

1. Experiment with support vector classification in Weka on the Iris2 dataset. Use the SMO algorithm. The complexity parameter is related to the complexity per iteration, keep it at 1. Choose to standardize the data, and keep the low order terms. What degree polynomial is needed to correctly classify the training set? Do the number of support vectors grow as the degree is increased?

Use RBF (radial basis functions). How does the number of support vectors and accuracy change as a function of the gamma parameter (which controls the width of the gaussians)?

2. Can linear support vector machines (without kernels) learn the XOR function? How about polynomial kernels?
3. Run AdaBoost on the Iris2 data in Weka. How many iterations does it take to get to training error zero?
4. Consider the following (unlabeled) sample where each instance contains three boolean features.

x_1	x_2	x_3
0	1	1
1	0	1
1	1	0
1	1	1
1	0	0
0	1	0

First, assume we have two distributions, $D_{2/3}$ and $D_{1/3}$, where D_p assigns probability $p^{x_1+x_2+x_3}(1-p)^{3-x_1-x_2-x_3}$ to the each instance (x_1, x_2, x_3) . In other words, each attribute x_i is given by independent flips of a 0-1 valued coin where p is the probability of 1. Calculate the probability of each instance under these two distributions. Calculate the likelihood of the sample under the mixture $(D_{2/3} + D_{1/3})/2$, corresponding to the model where nature first flips an unbiased coin to select which of the two D_p distributions to use, and then generates an instance from that distribution.

Second, *holding the distributions fixed* use EM to learn the mixture coefficients (coin bias) that maximize the likelihood of the sample. (I.e. for what $\alpha \in [0, 1]$ does distribution $\alpha D_{2/3} + (1 - \alpha) D_{1/3}$ maximize the probability of generating the sample? What is the likelihood of the sample under this optimal mixture? (You can run for 3 iterations rather than convergence if you prefer).

Third (optional): use EM to learn both the α values and the mixture proportions.

5. Consider the four-instance example used at the start of the boosting slides. Run AdaBoost (by hand) on that example to compute the weights of each example at each iteration (start with each example having weight $1/4$) and the weight of each of the three weak hypotheses. What is the final (un-normalized) margin of each of the four points?
6. Show that any second-order Markov process (where the new state depends on the previous two states) can be rewritten as a first-order Markov process.
7. Calculate the probability of observation sequence $(1,0,1,1,0,0)$ being generated, the sequence's Viterbi (most likely) path, and the probability of the Viterbi path in the following 2-state HMM with $\pi_1 = 1$ and $\pi_0 = 0$.

s	$P(s_1 s)$	$P(s_2 s)$	$P(1 s)$	$P(0 s)$
s_1	$3/4$	$1/4$	$3/4$	$1/4$
s_2	$1/4$	$3/4$	$1/3$	$2/3$