Data Mining II: Advanced Methods and Techniques

Assignment #2: Support Vector Machines

Please provide sufficient support so that your models could be reproduced and provide enough discussion to illustrate your understanding of the methods used and evaluate the model

1. What is a 'support vector'?
2. What is the "Kernel trick"?
3. What is the hyperplane and how is it utilized in SVMs?
4. Analyze the weather.arff data using Weka's Support Vector Machine method (Classifier.Functions.SMO).
    a) Use PolyKernel =1; How many support vectors were created? Would you say the produced model is acceptable? Was it able to successfully learn both outcomes of the 'play' class attribute?
    b) Change the parameter Kernel to "RBF Kernel: What has changed in the output? How many support vectors? Was the learning successful? Were both outcomes of 'play' successfully learned?
    c) Change the parameter Kernel to "Poly Kernel: = 2 or 3; Describe and explain the output? How many support vectors were created? Why was the number of kernel evaluations increased? Was the learning successful? Were both outcomes of 'play' successfully learned? How accurate is the model?
    d) What happens when you choose no normalization/standardization under the filter parameter?
    e) Which of the three models would you choose? Why?
5. Perform the same analysis on the Labor data set. Compare and contrast model/output between the two training data sets. Consider the class balance and size of the training data set.