

## CSE 6363 - *Machine Learning*

Homework/Project 2- Summer 2020

Due Date: July 27 2020, 7:00 pm Central Time (July 28, 8:00 Chinese Standard Time)

### Support Vector Machines

1. Consider the following linearly separable training data set:

$$D = \{ \begin{array}{l} ((1, 2), -1), \\ ((2, 3), 1), \\ ((2, 1), -1), \\ ((3, 4), 1), \\ ((1, 3), -1), \\ ((4, 4), 1) \end{array} \}$$

- a) Formulate the optimization function as well as the constraints for the corresponding linear maximum margin optimization problem without a regularization term. Also show the corresponding Lagrangian as well as the Lagrangian Dual for this problem.
- b) Use a SVM solver (e.g. MatLab's *fitsvm* function or Python's *sklearn.svm* from the scikit-learn toolbox) to learn the linear SVM parameters for this problem. Show the resulting decision boundary and identify the support vectors in this problem.

### Decision Trees

2. Consider the problem where we want to predict whether a balloon was inflated from a set of discrete attributes representing experiments with a person. The attributes are color (2 possible values), size (2 possible values), activity (2 possible values), and person's age (2 possible values). Data is given in the files as a comma separated list  $\{c, s, a, p, i\}$  where the first entry is the color (YELLOW or PURPLE), the second is the size (SMALL or LARGE), the third is the activity (STRETCH or DIP), the fourth entry is the person's age (CHILD or ADULT), and the last is the class to be predicted, i.e. whether it was inflated or not, (T or N). There is a training and a test data set for this problem.
  - a) Show the construction of a 2 level decision tree using minimum Entropy as the construction criterion on the training data set. You should include the entropy calculations and the construction decisions for each node you include in the 2-level tree.
  - b) Apply the tree from part a) to the test data set and compare the classification accuracy on this test set with the one on the training set. Does the result indicate overfitting ?