## CS392 Computer Learning : Exercises

### Learning outcomes

1. To revise the fundamentals of machine learning.
2. To revise the background theory of SVM's.
3. To look at small practical examples, and solve them using SVM's and least squares regression.
4. To discuss the theory of primal and dual forms.

### Questions

1. (a) Describe the general batch learning problem, what are its main components?
   (b) Describe using notation the format of a training set.
      i. What is a feature vector?
      ii. What are the labels in pattern recognition problems?
      iii. What are the labels in regression problems?
      iv. What is the task of the learning machine?
   (c) Describe how a medical diagnosis problem (such as predicting cancer risk) could be formulated as a pattern recognition problem. Hint: describe what the training and test examples would be, in terms of relevant features and labels.
   (d) Describe how we could predict the speed of a car, from its description by formulating the problem for regression analysis. Hint: again consider features and labels.

2. The following training set consists of noughts and crosses in the plane; the coordinates of the noughts are:

$$(0,-2),(0,-3),(1,-1),(1,-2),(1,-3),(2,1)$$

and the coordinates of the crosses are:

$$(0,1),(0,2),(-1,0),(-1,1),(-1,2),(-2,0),(-2,1)$$

(a) i. Find the separating hyperplane for this data.

ii. What does the separating hyperplane do?

iii. What are the features in this data, how many features $n$ are there?

iv. What are the labels $y_i$ in this example ?

v. How many training examples $l$ are there?

vi. What is the feature space?

(b) i. Which training examples are support vectors?

ii. In general what is the relationship between support vectors and the separating hyperplane?

(c) i. Find the predicted classifications ($\times$ or $\bigodot$) for the new points

$$(-1,3),(1,0),(1,-4)$$

ii. Which points can we be more confident in predicting and why?

3. (a) i. In support vector machine what is the equation for the separating hyperplane $H$?

   ii. Why do we try to find the optimal separating hyperplane for a set of training data?

   iii. What does the separating hyperplane represent?

   iv. What is the margin $\delta$ of a hyperplane.

   v. What are the support vectors?

   (b) Describe the difference between the separable and the non-separable cases of pattern recognition using support vector machines.

   (c) i. Define the optimisation problem, algebraically (eg. main optimisation and side constraints), for the SVM in the separable case.

   ii. What is the intuition behind minimising the weights?

   iii. What is the intuition behind the side constraints?

   (d) i. Define the optimisation problem, algebraically (eg. main optimisation and side constraints), for the SVM in the non-separable case.

   ii. What is the functionality of the slack variables?

4. You are given the training set,
   $x_1 = \begin{pmatrix} 1 & 0 & -1 & 0 & 0 & 1 & 1 \end{pmatrix}$, $y_1 = 0$ ,
   $x_2 = \begin{pmatrix} 0 & -1 & 0 & 1 & -1 & 0 & 1 \end{pmatrix}$, $y_2 = 1$
   and the new unlabelled test example
   $x_3 = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$

   (a) i. State the least-squares optimisation problem algebraically.
   ii. What is the matrix formulae for solving least-squares regression in primal form?
   iii. What is the matrix formulae for solving least-squares regression in dual form? Hint: take out the regularisation term from dual ridge regression in your notes.
   iv. In the dual form can we explicitly find the weight vector $\mathbf{w}$?

   (b) Find the least-squares prediction of the unseen label $y_3$ using either the primary or the dual variant of the least-squares method.

   (c) Justify your choice of method (i.e. primary or dual form). Could you have used the other method?

   (d) i. State the ridge regression optimisation problem algebraically.
   ii. How does this differ to the least-squares optimisation problem?
   iii. How could this modification improve the learning process cope with noisy data?