National University of Singapore
School of Computing
CS3244: Machine Learning
Tutorial 04-b

Regularization and Validation

1. **Cross Validation.** Consider a learning model that takes $m^2 \log m$ seconds of training when using m examples.

   (a) What is the total time (in seconds) needed when running leave one out cross validation(LOOCV) on 30 such models with different parameters to get the final hypothesis?

   (b) What is the total amount time (in seconds) needed when running 10-fold cross validation on 30 such models with different parameters to get the final hypothesis?

2. **Solving for Regularisation.** In this problem, we are working on a linear regression problem with regularisation on points in a 2D space. The figures 1, 2 below show contour plots for the linear regression problem with different regulariser. The ellipsis contours represent the squared error term. The circle and diamond contours represent the regularisation penalty term for $\lambda = 5$.

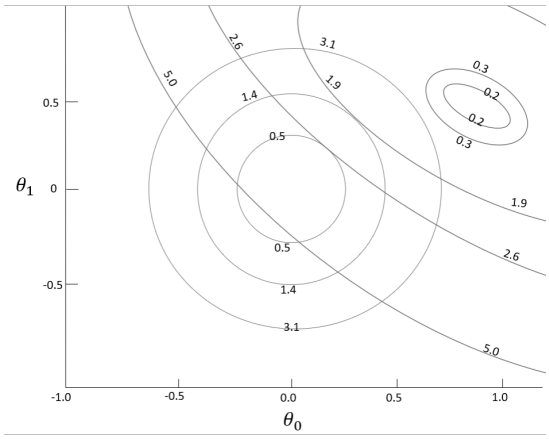

Figure 1: Contour plots for the linear regression problem with L2 regularisation

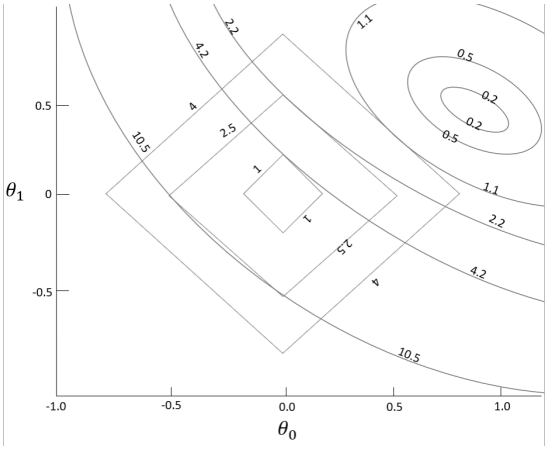

Figure 2: Contour plots for the linear regression problem with L1 regularisation

For each of the following cases, determine the rough values of θ_0 and θ_1 using the figures as reference.

(a) No regularisation.
(b) L2 regularisation with $\lambda = 5$.
(c) L1 regularisation with $\lambda = 5$.

3. **Variants of Validation.** In class, we have focused on validation on a set of data that is not time-based. Consider a recommender system where new shows are recommended for a consumer to watch based off their past selections and habits. Assume we have features of both the users and items. We also have time series data on the user's actions in such a case, in the form of tuples such as <user u watched show s at time t>.

   (a) Suppose we want to estimate the generalisation error of such a recommender system. What is a shortcoming of the usual validation technique used in class in this case?

   (b) Propose how you would modify the usual validation technique to fix this shortcoming.

   (c) Now we consider cross validation on time series data. Suppose we want to perform leave one out cross validation on this time series set of data [1, 2, 3, 4].

   Describe how you would perform cross validation, with which sets of data you would use to train and which set you would use as the testing data for each fold.