Assignment 1: k-nearest neighbours and linear regression

CS489/698 – Winter 2017

Out: January 9, 2016
Due: January 20 (11:59 pm), 2016.

Submit an electronic copy of your assignment via LEARN. Late submissions incur a 2% penalty for every rounded up hour past the deadline. For example, an assignment submitted 5 hours and 15 min late will receive a penalty of ceiling(5.25) * 2% = 12%.

Be sure to include your name and student number with your assignment.

1. **[30 pts]** Classification. Implement k-nearest neighbours using any programming language of your choice. Download the dataset posted on the course web page. Classify each input x according to the most frequent class amongst its k nearest neighbours. Break ties at random. Test the algorithms by 10-fold cross validation.

   What to hand in:

   - Your code for k-nearest neighbours and cross validation.
   - Find the best k by 10-fold cross validation. Draw a graph that shows the accuracy as k increases from 1 to 30.

2. **[30 pts]** Regression. Using any programming language, implement linear least square regression with the penalty term $\lambda w^T w$. Download the dataset posted on the course web page. The input and output spaces are continuous (i.e., $x \in \Re^d$ and $y \in \Re$).

   What to hand in:

   - Your code for linear regression.
   - Find the best λ by 10-fold cross validation. Draw a graph that shows the accuracy as λ increases from 0 to 4 in increments of 0.1.

3. **[40 pts]** Theory. In class, we discussed several loss functions for linear regression. However all the loss functions that we discussed assume that the error contributed by each data point have the same importance. Consider a scenario where we would like to give more weight to some data points. Our goal is to fit the data points (x_n, y_n) in proportion to their weights r_n by minimizing the following objective:

$$L(w, b) = \sum_{n=1}^{m} r_n (y_n - w x_n + b)^2$$

   where w and b are the model parameters, the training data pairs are (x_n, y_n). To simplify things, feel free to consider 1D data (i.e., x_n and w are scalars).

   (a) **[20 pts]** Derive a closed-form expression for the estimates of w and b that minimize the objective. Show the steps along the way, not just the final estimates.

   (b) **[20 pts]** Show that this objective is equivalent to the negative log-likelihood for linear regression where each data point may have a different Gaussian measurement noise. What is the variance of each measurement noise in this model?