

Week 2 Practical

Dr Simon Denman
CAB420: Machine Learning

This weeks practical will focus on regularised regression. It will build on the previous week's practical, and extending this model by adding higher order terms, and then seek to improve the model through regularisation.

The regularised regression lecture examples illustrate functions and/or packages that may be of use in completing this weeks practical.

Problem 1. Overfitting Linear Regression. In the week 1 practical, you developed a model to predict cyclist counts at a single counter. Using the same data and starting from the initial model before terms were removed, overcomplicate it such that it overfits to the training data. The easiest way to do this is by including a large number of higher order (i.e. interaction, quadratic and higher order polynomial) terms.

Verify that the model has overfit through evaluating on the validation and testing datasets, and compare it's performance to the simple model that you started with.

Problem 2. Ridge Regression. Apply ridge regression to your two models (the simple model from Week 1, and the overfitting model from Problem 1 of this week). Using the validation set, select the best value of λ for each model. For the selected model:

- Compute the R^2 and adjusted R^2 , and draw a qqplot to assess the models validity;
- Compute the RMSE on the test set and compare the performance with the linear models.

Problem 3. Lasso Regression. Apply lasso regression to your two models (the simple model from Week 1, and the overfitting model from Problem 1 of this week). Using the validation set, select the best value of λ for each model. For the selected model:

- Compute the R^2 and adjusted R^2 (make sure to consider how many terms are removed by lasso), and draw a qqplot to assess the models validity;
- Compute the RMSE on the test set and compare the performance with the linear models and the ridge regression models.