Computer practical: Prediction for Machine Learning

Sebastian Funk, Jeremy Brown, Stephen O'Neill

2022-01-25

Sebastian Funk, Jeremy Brown, Stephen O'Neill

EXERCISE

- Make sure you have a trained model with predictions from Monday
- Calculate the sensitivity, specificity, PPV, NPV, etc. for your trained model for both the training and test sets
- Plot the ROC for your classifier (test set)
- How does your classifier compare to random label allocation?
- How confident are you in your predicted positives?
- How confident are you in your predicted negatives?
- Under what situation would this classifier be useful?

continued…

Sebastian Funk, Jeremy Brown, Stephen O'Neill

EXERCISE (continued)

- Calculate Youden's J to find the optimal threshold
- Calculate new class labels using Youden's J
- What are the new sensitivity, specificity, PPV and NPV?
- Under what situation would this classifier be useful?

Sebastian Funk, Jeremy Brown, Stephen O'Neill

EXTENSION EXERCISE (optional)

- Fit another model and compare the two using the log score and the Brier score.
- hint: for the log score you can use the convenient forula

$$L = -\frac{1}{N} y_i \log p_i + (1 - y_i) \log(1 - p_i)$$

- Do the different scores give the same result as e.g. using misclassification error in terms of which of the models do best?
- Test whether the model would make better predictions with a more data to train it: Re-train the models with a larger training set and predict with the same test set
- Compare your model prediction results

Sebastian Funk, Jeremy Brown, Stephen O'Neill

EXTENSION EXERCISE (Optional)

- Re-train the model on a training set which is selected not completely at random, e.g. consider training on examples where BMI is less than 25
  - Hint: keep the same test set but make sure none of the observations in it end up in the training set
- Re-test the model on your test set and discuss predictions for observed presences

Sebastian Funk, Jeremy Brown, Stephen O'Neill

EXTENSION EXERCISE (Optional)

- Revisit your week 1 regression models of systolic blood pressure and choose one regression model approach
- Create a training and testing set as you did above
- What is the R^2 value for this model for both in- and out of sample predictions?
- What is the RMSE for this model for both in- and out of sample predictions?
- Calculate the RMSE for in- and out of sample predictions from a model which is just the mean
- How does your regression model compare to just taking the mean?
- Under what situation would this regression model be useful?

Computer practical: Prediction for Machine Learning

Sebastian Funk,
Jeremy Brown,
Stephen O'Neill