Chapter 1.3 - Model Selection

Paul Scemama

September 2022

1 Model Selection

We saw in 1.1 that the choice of M affected performance, as well as regularization levels. In practice we need a way to determine the value M.

- As well as finding values for *complexity parameters*, we may wish to consider a range of different model types.

We've also seen that when using maximum likelihood, training performance is not always a good indicator of test performance due to overfitting.

- If data is plentiful, we can compare various models via a hold-out set called a *validation set.*
- If our iterative model design is done using limited data, some overfitting to the validation set can occur; in which case we could have a third test set on which performance of the selected model is finally evaluated.

We want to make the most out of our data for training.

- This can lead to a small validation set $\rightarrow$ noisy estimates of performance.
- *solution*: Cross-validation
  - proportion $\frac{(S-1)}{S}$ used for training, then test on remaining $\frac{1}{S}$.
  - Do this S times.

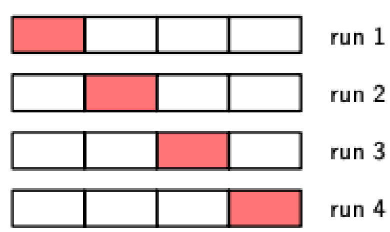

Figure 1.18 The technique of S-fold cross-validation, illustrated here for the case of $S = 4$, involves taking the available data and partitioning it into S groups (in the simplest case these are of equal size). Then $S - 1$ of the groups are used to train a set of models that are then evaluated on the remaining group. This procedure is then repeated for all S possible choices for the held-out group, indicated here by the red blocks, and the performance scores from the S runs are then averaged.

  –
  - *drawbacks*: the number of training runs that must be performed increases by a factor of S; and on another note, we may have multiple complexity parameters to estimate per model.

Ideally we want a technique that allows multiple hyper-parameters and model types to be compared in a single training run.

- We therefore need a measure of performance which depends only on the training data and which does not suffer from bias due to overfitting

- Various *information* criterion have been proposed to try and correct for the bias of maximum-likelihood.
  - Examples: *AIC* and *BIC*
  - Such criteria do not take into account uncertainty in model parameters; however in practice they tend to favor overly simple models.

We will see later that a fully Bayesian approach allows for complexity penalties to arise in a natural way.