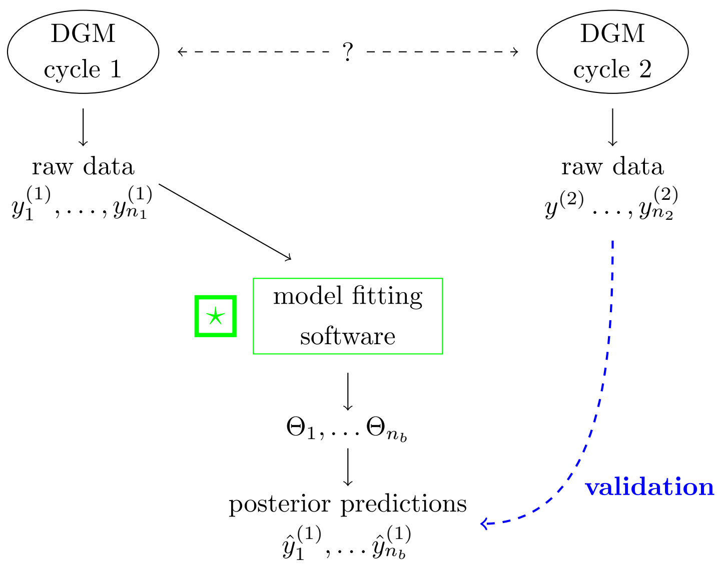
A Flowchart for Prediction Scoring

Anna Smith

August 28, 2017

1 Tian's flowchart



`score(` $\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, \boxed{\star}$, ...)

2 Validation and cross validation

cross validation:
for each $k = 1, \ldots, n_1$,

$$\boldsymbol{y}_{-k}^{(1)} \longrightarrow \boxed{\star} \xrightarrow{\ \boldsymbol{x}_k^{(1)}\ } \hat{\boldsymbol{y}}_k^{(1)},$$

and compare $y_k^{(1)}$ to $\hat{\boldsymbol{y}}_k^{(1)}$ by computing a quantile,

$$q_k^{(1)} = \frac{1}{L} \sum_{l=1}^{L} I_{\left\{ y_k^{(1)} > \hat{y}_{kl}^{(1)} \right\}}.$$

That is, we fit the model to $\boldsymbol{y}_{-k}^{(1)}$, the original dataset with the kth observation removed, and then use $\boldsymbol{x}_k^{(1)}$, any covariates corresponding to the kth observation, and this fitted model to produce a vector of predictions for the kth observation, $\hat{\boldsymbol{y}}_k^{(1)}$. If the model fits the data well, then the true value, $y_k^{(1)}$, should look like it belongs with this vector of predictions, $\hat{\boldsymbol{y}}_k^{(1)}$, and the $q_k^{(1)}$ should be uniformly distributed.

validation:
for each $k = 1, \ldots, n_2$,

$$\boldsymbol{y}^{(1)} \longrightarrow \boxed{\star} \xrightarrow{\ \boldsymbol{x}_k^{(2)}\ } \hat{\boldsymbol{y}}_k^{(2)},$$

and compare $y_k^{(2)}$ to $\hat{\boldsymbol{y}}_k^{(2)}$ by computing a quantile,

$$q_k^{(2)} = \frac{1}{L} \sum_{l=1}^{L} I_{\left\{ y_k^{(2)} > \hat{y}_{kl}^{(2)} \right\}}.$$

That is, we fit the model to $\boldsymbol{y}^{(1)}$, the full cycle one dataset, and then use $\boldsymbol{x}_k^{(2)}$, any covariates corresponding to the kth observation from cycle two, and this fitted model to produce a vector of predictions for the kth observation in cycle two, $\hat{\boldsymbol{y}}_k^{(2)}$. If the model fits the data well, then the true value, $y_k^{(2)}$, should look like it belongs with this vector of predictions, $\hat{\boldsymbol{y}}_k^{(2)}$, and the $q_k^{(2)}$ should be uniformly distributed.

3 Prediction scoring

It may be tempting to stop here and treat the set of $q_k^{(2)}$'s as prediction scores, by considering how closely they follow the uniform distribution, for example, or by deriving a more nuanced test statistic from this set of measures. However, any tests about the $q_k^{(2)}$'s include the assumption that our model is correct. Indications from the $q_k^{(2)}$'s alone that the data generating mechanism varies across the two cycles could simply be due to an inadequate model.

Instead, we recommend comparing the $q_k^{(2)}$'s to the $q_k^{(1)}$'s. Because both sets of quantiles are based on the same model, comparisons across these vectors should be less sensitive to poor modelling choices.