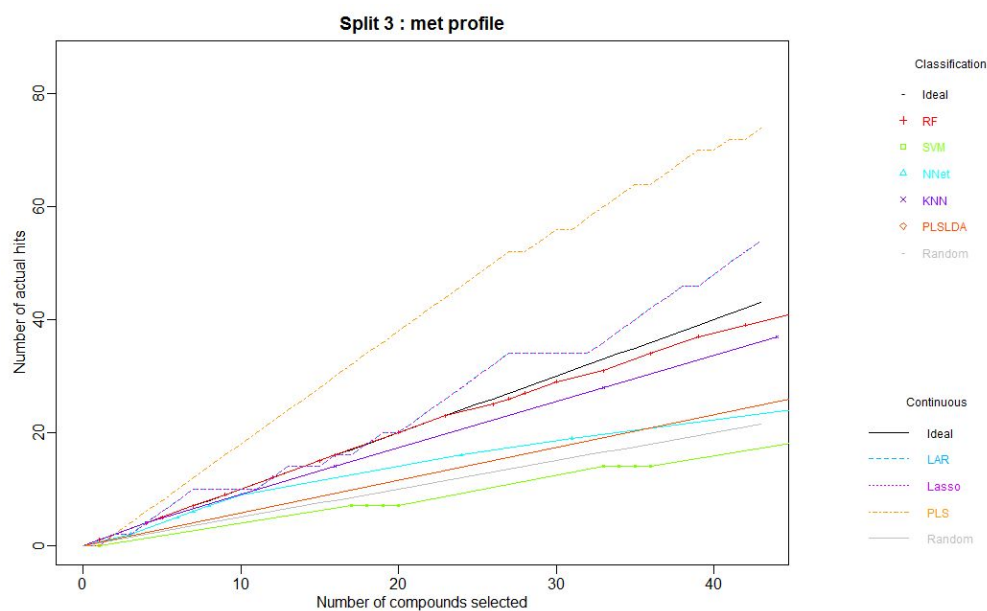


# Changes before submission to CRAN

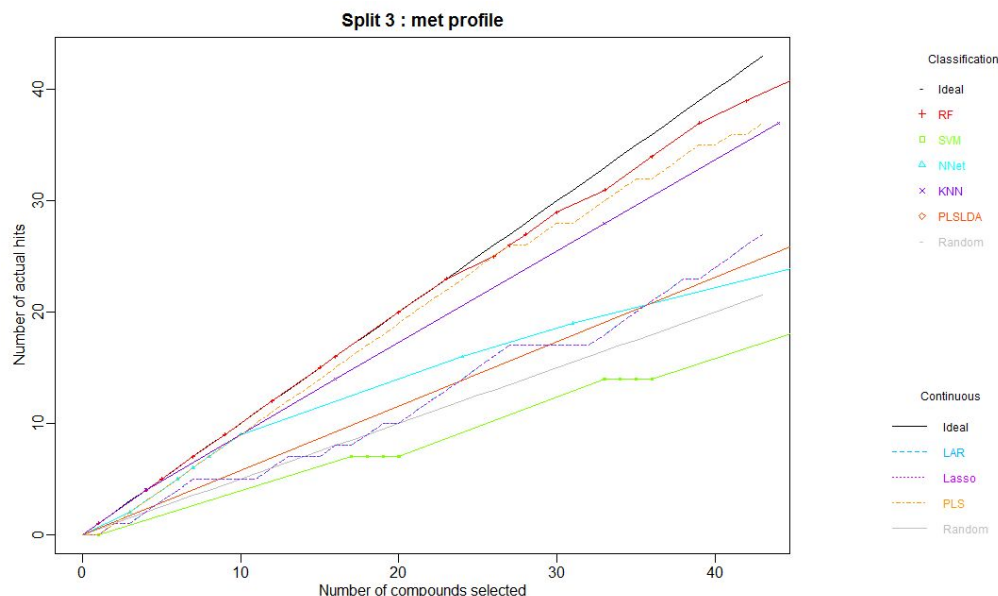
## Issues to resolve

- Analyzed a data set where hit rate was higher than ideal curve - SOLVED (really)
  - The scale of the y-axis was different for continuous and classification curves when they were plotted together on the same plot.
  - The maximum y value was set to the total number of actives for classification curves. For continuous curves, the maximum y value was set to the number of actives in the set of compounds selected for plotting (selected by max.select parameter). If the max.select value is less than the number of actives, then the y axis for continuous and classification curves were on a different scale.
  - Figure showing original error:



- 
- 
-

- Figure showing the error resolved:



- 
- The methods LAR and LASSO seem to give exactly the same output. I know at one point we talked about using slightly different implementations, and I can't remember where we landed. Are we intentionally getting the same output for LAR and LASSO? Or is this just by happenstance with the two datasets I've tested?
  - I have noticed that same issue. I am using the lar package for both LAR and Lasso. The Lasso solution path is produced by a slight modification of the same LAR algorithm. I noticed that in the original chemmodlab LAR implementation, you were creating the solution path and then predicting on each held out fold using the last "step" of the solution path. I have checked the lars output and this is the step where the ratio of the L1 norm of the coefficient vector, relative to the norm at the full LS solution, is 1. This is simply linear regression without a penalty. I used the same approach with Lasso. This resulted in the same linear regression fit.
  - Additionally, I have noticed that in a simple test case, The whole solution path was the same for the lar and lasso algorithms:
  - Lasso solution path (notice the last step is the model with all predictors):
    - `attach(diabetes)`
    - `> object <- lars(x,y, type = "lasso")`
    - `> plot(object)`

- 
- Lars solution path (notice the last step is the model with all predictors):
  - `> object <- lars(x,y, type = "lar")`
  - `> plot(object`

- 
- To address this, I have switched to setting the lasso lambda penalty parameter to .2 by default. This results in a model with many less predictors in the few test cases I looked at. This will enable users to tune and customize the penalty parameter for lasso.
- I also set made the lar lambda parameter tunable, but set the lambda value to 0.05 as a default. This will result in one more sparse one less sparse model. These seemed sensible. Since lar and lasso are giving similar solution paths we

might as well look at two different points on the solution path and not provide near identical solutions by default. I don't know what are good defaults here...

- Now LAR and Lasso have different performance. But the lambda values I chose don't have great performance.

○

- The MCS plot currently has a label of "Model Accuracy". This is misleading the user because it is not always "accuracy" - DONE
  - Changed to "Model Performance" (That was the best I could get the bottom left aligned)

○

- The dotted accumulation curve (dots only, not dot-dash) is extremely hard to see.
  - Changed lwd from 1 to 2 for continuous curves. I think I liked the previous setting better before. Cannot find a way to set line width to an intermediate value at the moment.:

- 
- Is "at" activated for ppv/auc/fmeasure in function Performance? It needs to be -DONE
  - Updated the documentation, this was already implemented for everything but auc. Implemented for auc now too.

## New features to implement

- Add performance measures - DONE
  - PPV
  - F1-measure
- Use at argument for sensitivity, specificity, error rate - DONE
  - See CombineSplits/Performance documentation
- Multiple model performance measures output by Performance function - DONE

## Other Things TODO

- Check all ways of specifying descriptor sets and responses is working correctly - DONE
- Check all performance measures - DONE
- Discontinue LassoGLM - DONE
- Remove vignette - DONE
  - Overlap with R journal paper
- 

## Documentation

- Document the aid364 data set - DONE