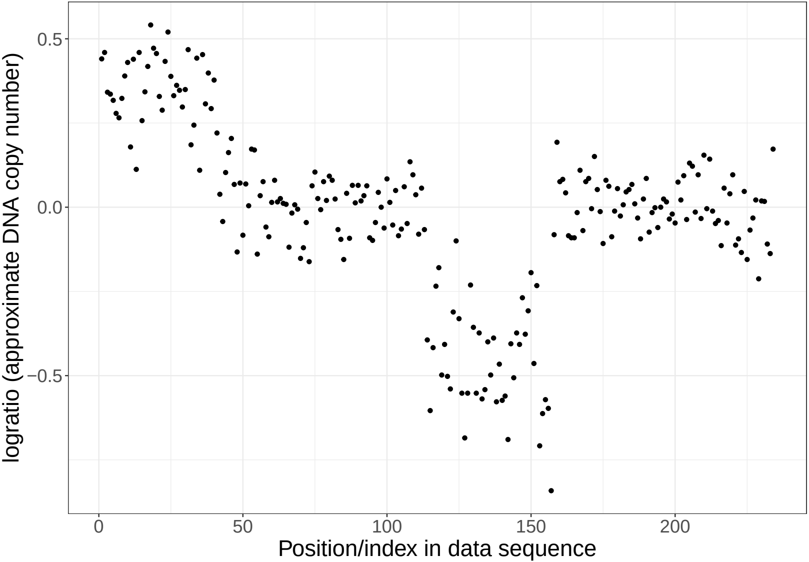
Segmentation model selection and evaluation

Toby Dylan Hocking

Background: detecting abrupt changes is important

Example from cancer diagnosis: breakpoints are associated with aggressive disease in neuroblastoma.



Motivation for segmentation model selection and evaluation

- In each of the segmentation models we have studied, there is a choice of model size (segments/changepoints or hidden states).
- Too large model sizes result in false positives (changepoints predicted by algorithm but they are not significant/real).
- Too small model sizes result in false negatives (no changepoint predicted where there should be).
- Want to maximize true positive rate (number of correctly predicted changepoints) and true negative rate (number of correct predicted regions without changepoints).

Model selection via Classic Information Criteria

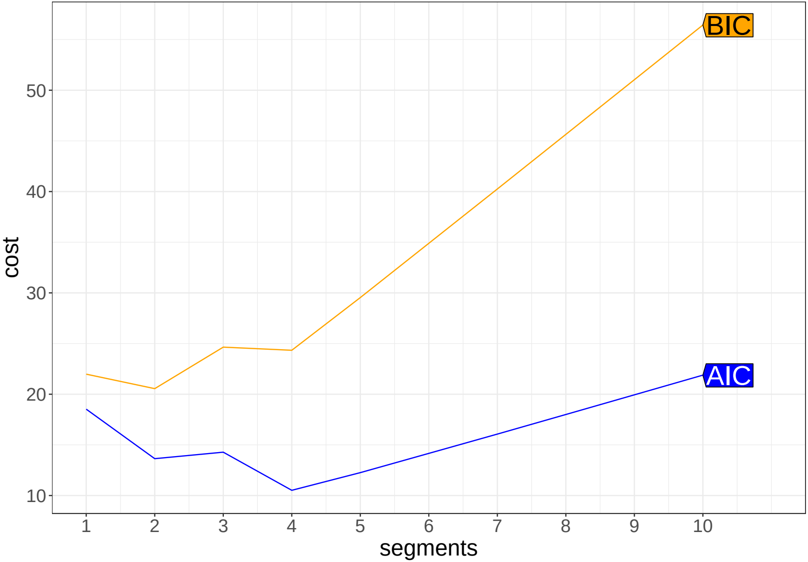
For every model size $k \in \{1, \dots, K\}$ let L_k be the loss.

Information criteria choose the model which minimizes the penalized cost, for some non-negative penalty $\lambda \geq 0$,

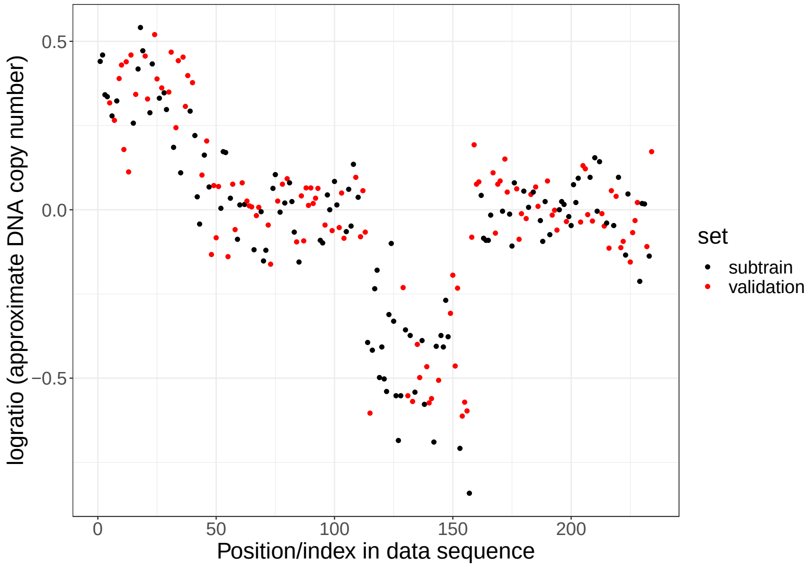
$$C(\lambda) = \min_{k \in \{1,\dots,K\}} L_k + \lambda k$$

- BIC=Bayesian Information Criterion, sometimes referred to as SIC=Schwarz who was the author. $\lambda = \log n$ where n is the number of data points.
- AIC=Akaike Information Criterion: $\lambda = 2$.
- Data viz: http://bl.ocks.org/tdhock/raw/43ac9c6be9188dcb02a7/

Model selection criteria plot for binary segmentation



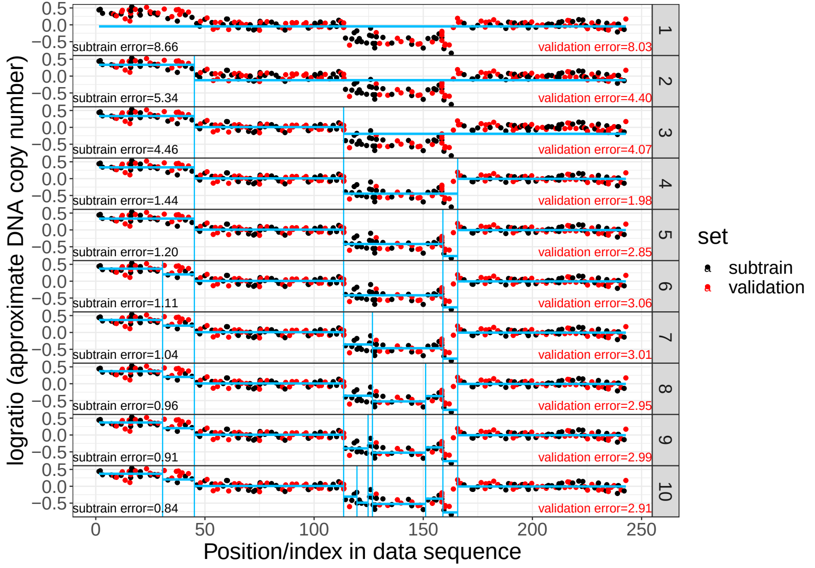
Cross-validation for model selection



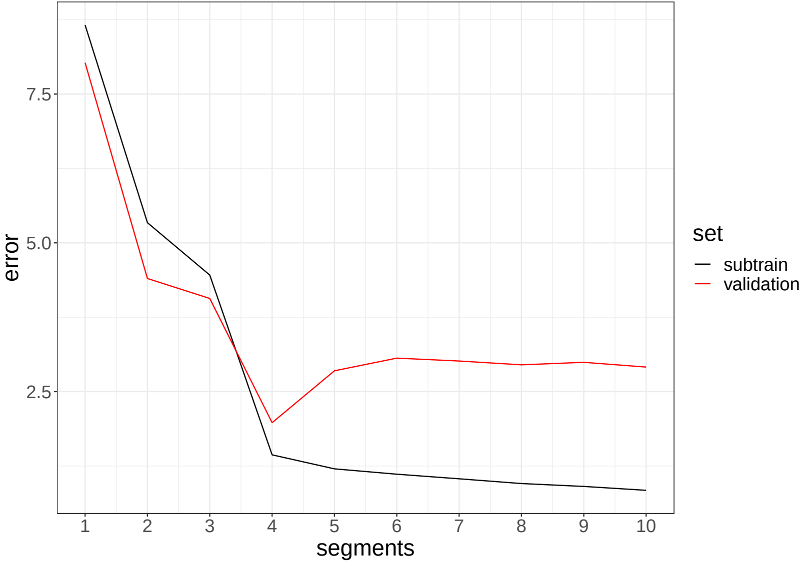
Idea for cross-validation

- Divide full data sequence into subtrain and validation sets.
- Use subtrain data as input to learning algorithm.
- Compute predicted changepoint positions and segment parameters using only subtrain data.
- Assign parameters to validation data based on predicted changepoints.
- Use validation data to choose best model size (min error or negative log likelihood).
- As model size increases, subtrain error should always decrease, whereas validation error should be U shaped.

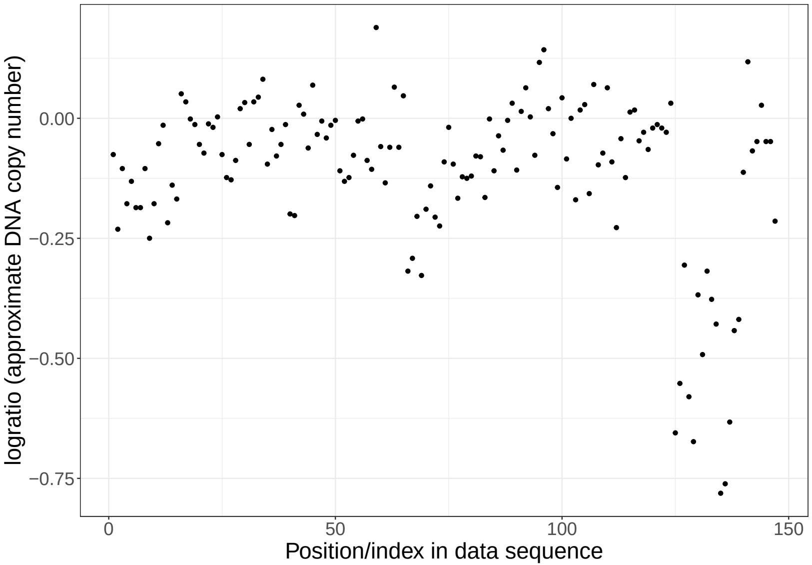
Fitting model to validation set



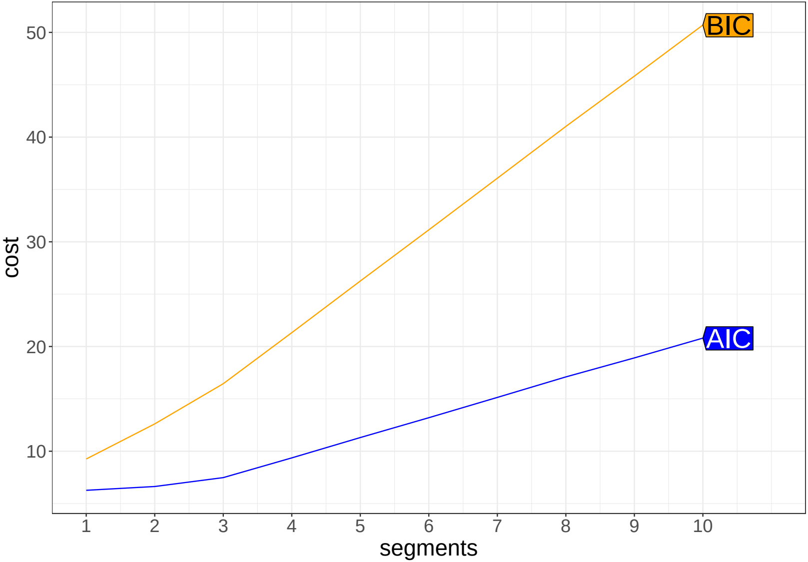
CV Error plot



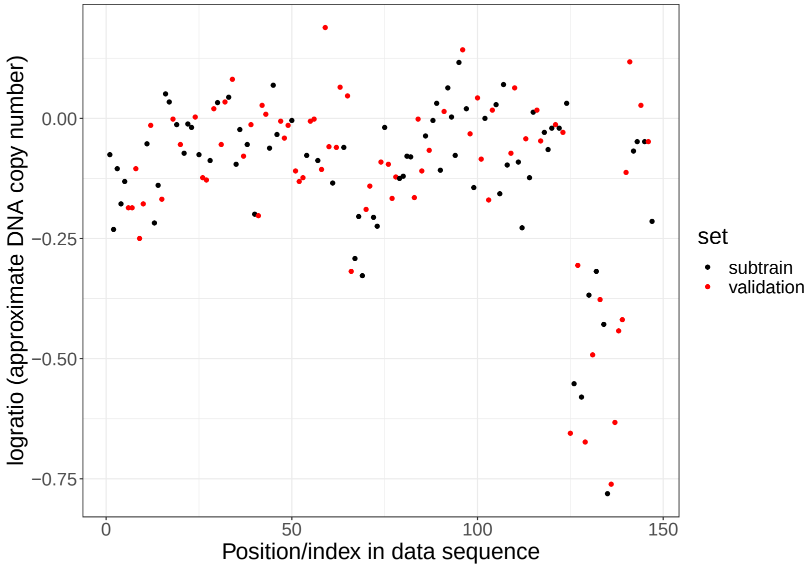
Another data set



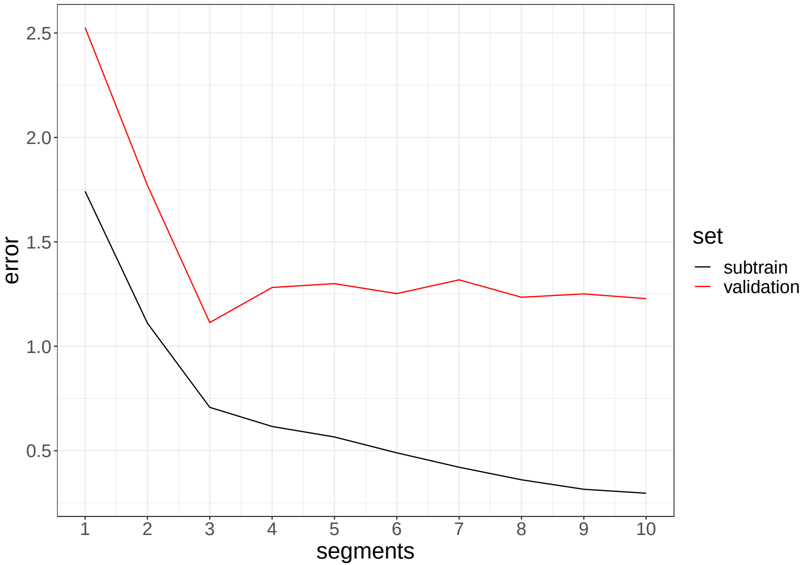
Model selection plot



Cross-validation for model selection



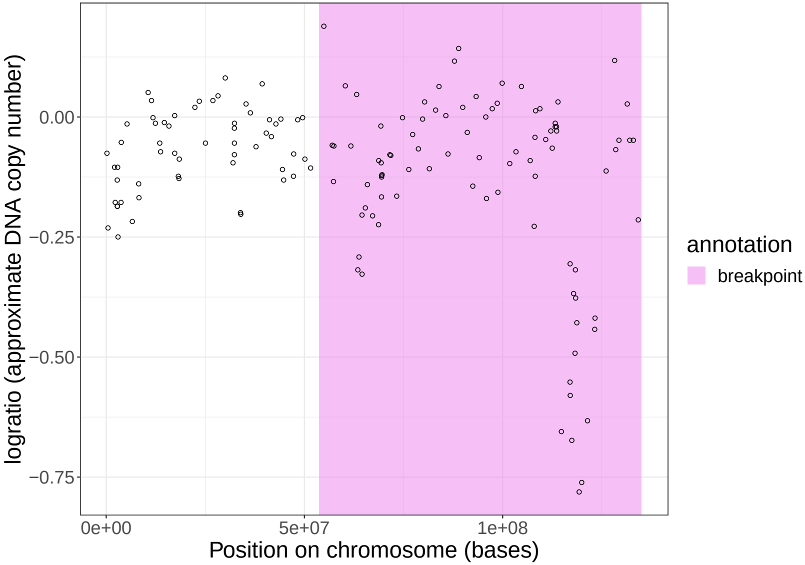
CV error plot



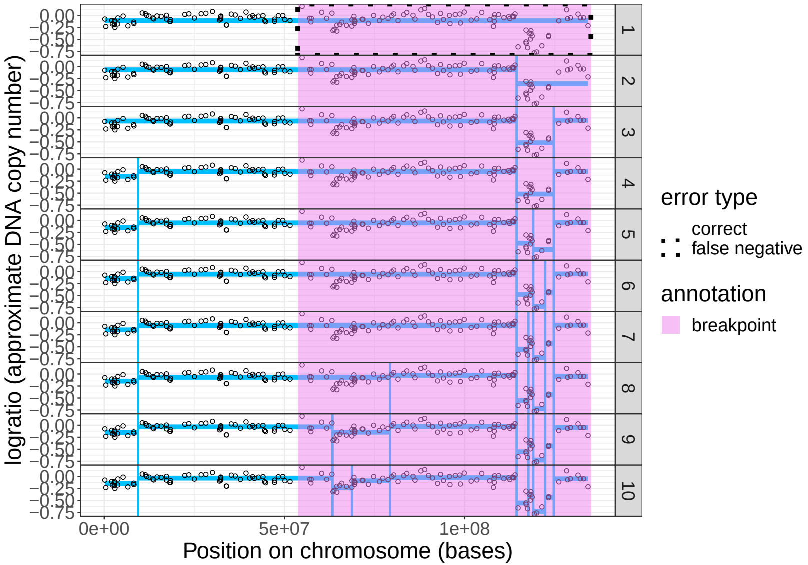
Labeled regions for evaluating accuracy of changepoint predictions

- After we have selected one penalty/model, how to quantify its error rate? (false positive, false negative)
- In general with real data, this is a difficult/unsolved problem.
- Sometimes labels can be created (prior knowledge, visual inspection).
- Like in unsupervised clustering the labels are assumed to be only available after the model has been learned.

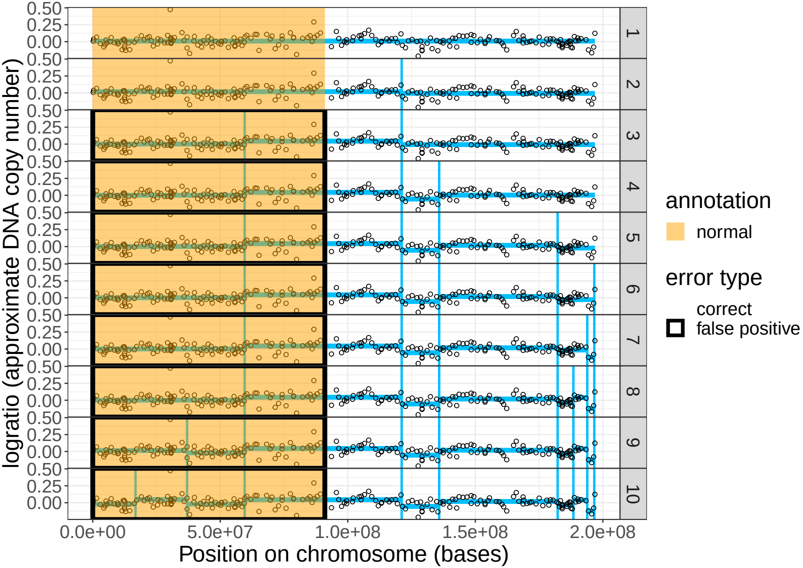
Labeled regions for evaluating accuracy of changepoint predictions



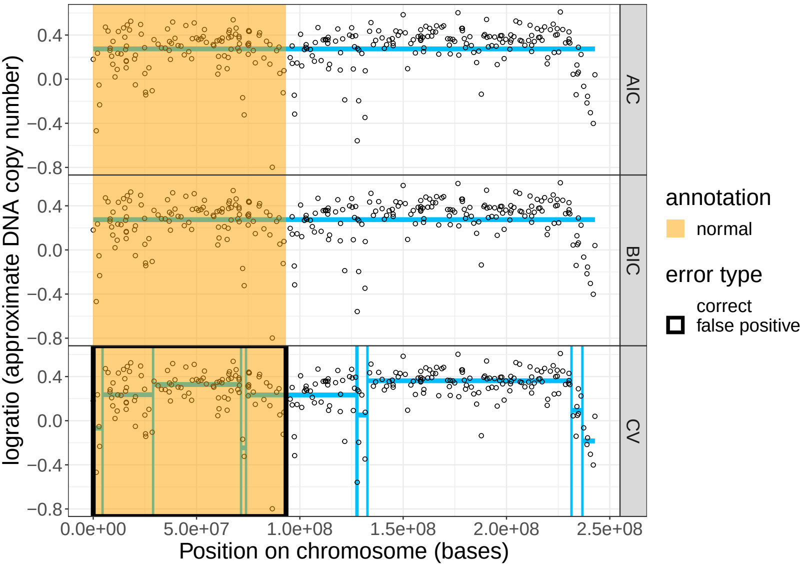
False negative for missing change in positive label



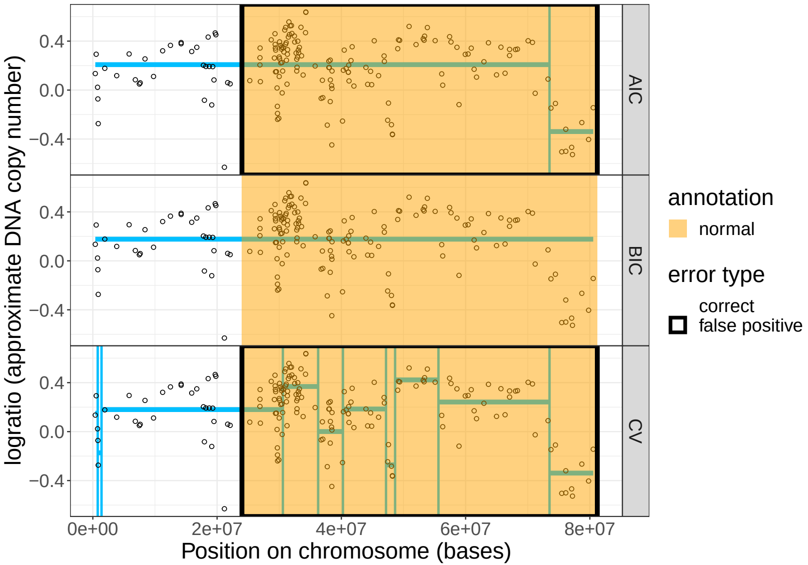
False positive for changepoint in negative label



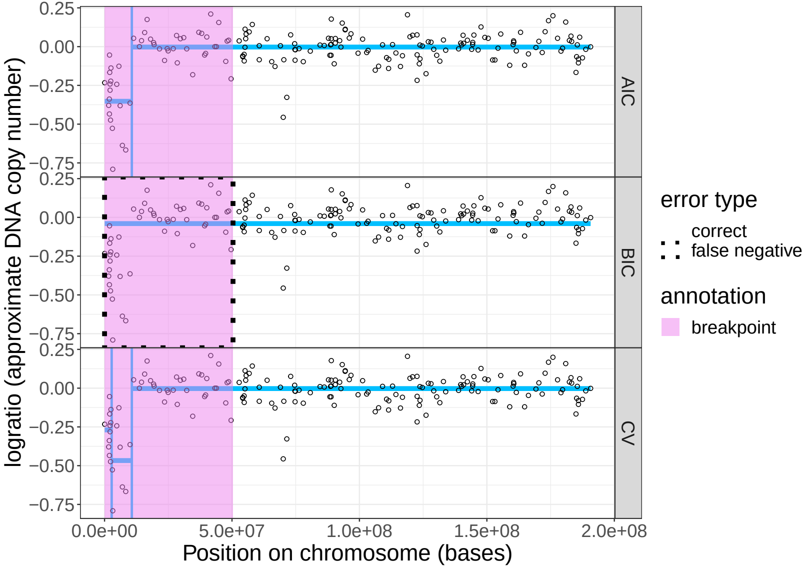
Comparing model selection algorithms



Comparing model selection algorithms

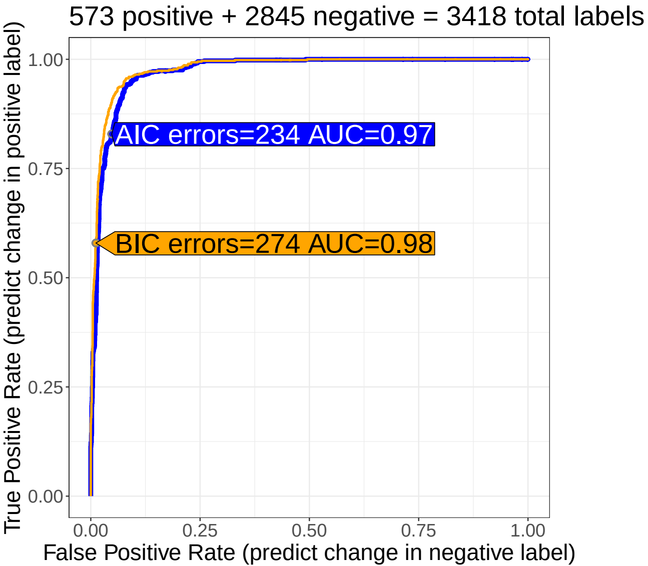


Comparing model selection algorithms



ROC curves for evaluation

- Point shown is error rate of predicted penalty.
- Add constants to that penalty to trace ROC curve.



Other ROC curve data visualizations

- http://ml.nau.edu/viz/2021-10-21-curveAlignment/
- http://ml.nau.edu/viz/2021-10-21-neuroblastomaProcessed-complex/

Possible exam questions

- What is the difference between AIC and BIC?
- Does AIC or BIC tend to select larger model sizes? or do they select the same model size? why?
- When using cross-validation for model selection, what set is used as input to the learning algorithm? What set is not input to the learning algorithm?
- What modifications would be required to use K-fold cross-validation instead of a single 50% subtrain, 50% validation split?
- What modifications, if any, would be needed to use these model selection algorithms with Hidden Markov Models?