Prediction Accuracy Metrics

Goals for Today

- To understand how to compute
  - MAE -- Mean absolute error
  - MSE -- Mean squared error
  - RMSE -- Root mean squared error
- To understand variations on how these may be computed
- To understand where accuracy metrics are useful in general, and the relative merits of each of these three

A little intuition

- Error metrics are usually computed using a "leave one out" methodology
  - Cover up a rating, and try to predict it
- Warning: sometimes this is hard, and evaluators take short cuts (e.g., leave 10% out).

Mean Absolute Error (MAE)

- What is error?
  - Divergence of prediction from actual opinion (rating)
  - P-R
- Absolute error removes direction
  - | P – R |
  - Why? Because two wrongs don't make a right!
- MAE = Average (| P – R |)
  - $\frac{\sum_{ratings} |P-R|}{\#\ ratings}$

Mean Squared Error (MSE)

- Why Squared Error?
  - Removes sign – avoids need for absolute value
  - Penalizes large errors more than small
- $\frac{\sum_{ratings}(P-R)^2}{\#\ ratings}$
- One disadvantage – squared error is not on an intuitive scale …

Root Mean Squared Error (RMSE)

- $\sqrt{\frac{\sum_{ratings}(P-R)^2}{\#\ ratings}}$

Hold on a moment …

- We glossed over the summation
  - Usual model – average over all ratings
  - Alternative model – average over user averages
- What's the difference
  - What if one user has 3000 ratings and another 10?
- Advice – consider looking at both – understand what you're comparing to

Comparing Different Algorithms

- What to do when computing MAE in different cases:
  - Remember, must be same data set/scale
  - If coverage is different (different set of user/item pairs for which predictions are available, two choices):
    - Check against common subset
    - Supplement algorithm with default for full coverage

Reflections …

- In general, all the error metrics move together (good replacements for each other)
- Squared may matter for large scales with some algorithms that have occasional huge errors, but other measures may catch that better
- Benefit – lots of published MAE, RMSE data for public datasets
- Drawback – error can be dominated by irrelevant parts of the item space

Looking forward …

- Next, we look at decision-support metrics

Prediction Accuracy Metrics