eval\_results\_autotag\_M\_20\_1\_fold\_1\_times\_all\_attempts.RData

> load(f)

> show.eval(eval.results)

Number of observations used: 172410

M = 20 eta = 0

-LL = 0.426 (overall and problem-specific learned chance would give 0.366 and 0.207 respectively)

-LL correct = 0.314 (overall and problem-specific learned chance would give 0.166 and 0.117 respectively)

-LL incorrect = 0.857 (overall and problem-specific learned chance would give 1.142 and 0.555 respectively)

MAE = 0.273 (overall and problem-specific learned chance would give 0.326 and 0.19 respectively)

RMSE = 0.428 (overall and problem-specific learned chance would give 0.404 and 0.308 respectively)

[1] "Confusion matrix:"

Incorrect Correct

Predict Incorrect 10.0 13.9

Predict Correct 10.6 65.6

[1] "True: 75.6"

[1] "Confusion matrix of overall chance:"

Incorrect Correct

Predict Incorrect 0.0 0.0

Predict Correct 20.5 79.5

[1] "True: 79.5"

[1] "Confusion matrix of specific chance:"

Incorrect Correct

Predict Incorrect 11.3 5.3

Predict Correct 9.2 74.2

[1] "True: 85.5"

eval\_results\_autotag\_M\_20\_1\_fold\_1\_times\_first\_attempts.RData

> load(f)

> show.eval(eval.results)

Number of observations used: 87735

M = 20 eta = 0

-LL = 0.588 (overall and problem-specific learned chance would give 0.376 and 0.249 respectively)

-LL correct = 0.514 (overall and problem-specific learned chance would give 0.175 and 0.138 respectively)

-LL incorrect = 0.859 (overall and problem-specific learned chance would give 1.107 and 0.651 respectively)

MAE = 0.382 (overall and problem-specific learned chance would give 0.338 and 0.23 respectively)

RMSE = 0.514 (overall and problem-specific learned chance would give 0.411 and 0.339 respectively)

[1] "Confusion matrix:"

Incorrect Correct

Predict Incorrect 10.3 24.8

Predict Correct 11.2 53.7

[1] "True: 64"

[1] "Confusion matrix of overall chance:"

Incorrect Correct

Predict Incorrect 0.0 0.0

Predict Correct 21.5 78.5

[1] "True: 78.5"

[1] "Confusion matrix of specific chance:"

Incorrect Correct

Predict Incorrect 8.6 4.5

Predict Correct 12.9 73.9

[1] "True: 82.6"

eval\_results\_autotag\_M\_20\_1\_fold\_1\_times\_last\_attempts.RData

> load(f)

> show.eval(eval.results)

Number of observations used: 87735

M = 20 eta = 0

-LL = 0.504 (overall and problem-specific learned chance would give 0.231 and 0.159 respectively)

-LL correct = 0.441 (overall and problem-specific learned chance would give 0.074 and 0.063 respectively)

-LL incorrect = 1.085 (overall and problem-specific learned chance would give 1.677 and 1.048 respectively)

MAE = 0.315 (overall and problem-specific learned chance would give 0.176 and 0.133 respectively)

RMSE = 0.475 (overall and problem-specific learned chance would give 0.297 and 0.258 respectively)

[1] "Confusion matrix:"

Incorrect Correct

Predict Incorrect 3.6 23.4

Predict Correct 6.2 66.8

[1] "True: 70.4"

[1] "Confusion matrix of overall chance:"

Incorrect Correct

Predict Incorrect 0.0 0.0

Predict Correct 9.8 90.2

[1] "True: 90.2"

[1] "Confusion matrix of specific chance:"

Incorrect Correct

Predict Incorrect 1.5 0.7

Predict Correct 8.3 89.5

[1] "True: 91"

eval\_results\_autotag\_M\_20\_3\_fold\_10\_times\_all\_attempts.RData

> load(f)

> show.eval(eval.results)

Number of observations used: 1724100

M = 20 eta = 0

-LL = 0.439 (overall and problem-specific learned chance would give 0.366 and 0.209 respectively)

-LL correct = 0.326 (overall and problem-specific learned chance would give 0.166 and 0.118 respectively)

-LL incorrect = 0.876 (overall and problem-specific learned chance would give 1.142 and 0.56 respectively)

MAE = 0.275 (overall and problem-specific learned chance would give 0.326 and 0.19 respectively)

RMSE = 0.435 (overall and problem-specific learned chance would give 0.404 and 0.309 respectively)

[1] "Confusion matrix:"

Incorrect Correct

Predict Incorrect 9.8 14.7

Predict Correct 10.7 64.8

[1] "True: 74.6"

[1] "Confusion matrix of overall chance:"

Incorrect Correct

Predict Incorrect 0.0 0.0

Predict Correct 20.5 79.5

[1] "True: 79.5"

[1] "Confusion matrix of specific chance:"

Incorrect Correct

Predict Incorrect 11.4 5.5

Predict Correct 9.1 74.0

[1] "True: 85.4"

eval\_results\_autotag\_M\_20\_3\_fold\_10\_times\_first\_attempts.RData

> load(f)

> show.eval(eval.results)

Number of observations used: 877350

M = 20 eta = 0

-LL = 0.596 (overall and problem-specific learned chance would give 0.376 and 0.251 respectively)

-LL correct = 0.512 (overall and problem-specific learned chance would give 0.175 and 0.139 respectively)

-LL incorrect = 0.903 (overall and problem-specific learned chance would give 1.107 and 0.662 respectively)

MAE = 0.383 (overall and problem-specific learned chance would give 0.338 and 0.23 respectively)

RMSE = 0.515 (overall and problem-specific learned chance would give 0.411 and 0.34 respectively)

[1] "Confusion matrix:"

Incorrect Correct

Predict Incorrect 10.2 24.6

Predict Correct 11.4 53.9

[1] "True: 64"

[1] "Confusion matrix of overall chance:"

Incorrect Correct

Predict Incorrect 0.0 0.0

Predict Correct 21.5 78.5

[1] "True: 78.5"

[1] "Confusion matrix of specific chance:"

Incorrect Correct

Predict Incorrect 8.7 4.7

Predict Correct 12.8 73.7

[1] "True: 82.5"

eval\_results\_autotag\_M\_20\_3\_fold\_10\_times\_last\_attempts.RData

> load(f)

> show.eval(eval.results)

Number of observations used: 877350

M = 20 eta = 0

-LL = 0.522 (overall and problem-specific learned chance would give 0.231 and 0.162 respectively)

-LL correct = 0.438 (overall and problem-specific learned chance would give 0.074 and 0.063 respectively)

-LL incorrect = 1.302 (overall and problem-specific learned chance would give 1.678 and 1.074 respectively)

MAE = 0.314 (overall and problem-specific learned chance would give 0.176 and 0.134 respectively)

RMSE = 0.479 (overall and problem-specific learned chance would give 0.297 and 0.259 respectively)

[1] "Confusion matrix:"

Incorrect Correct

Predict Incorrect 3.5 23.5

Predict Correct 6.3 66.8

[1] "True: 70.2"

[1] "Confusion matrix of overall chance:"

Incorrect Correct

Predict Incorrect 0.0 0.0

Predict Correct 9.8 90.2

[1] "True: 90.2"

[1] "Confusion matrix of specific chance:"

Incorrect Correct

Predict Incorrect 1.6 0.8

Predict Correct 8.2 89.4

[1] "True: 90.9"

eval\_results\_SMEtag\_M\_20\_1\_fold\_1\_times\_all\_attempts.RData

> load(f)

> show.eval(eval.results)

Number of observations used: 56178

M = 20 eta = 0

-LL = 0.419 (overall and problem-specific learned chance would give 0.476 and 0.298 respectively)

-LL correct = 0.424 (overall and problem-specific learned chance would give 0.335 and 0.244 respectively)

-LL incorrect = 0.41 (overall and problem-specific learned chance would give 0.714 and 0.391 respectively)

MAE = 0.329 (overall and problem-specific learned chance would give 0.467 and 0.282 respectively)

RMSE = 0.444 (overall and problem-specific learned chance would give 0.483 and 0.376 respectively)

[1] "Confusion matrix:"

Incorrect Correct

Predict Incorrect 28.6 22.0

Predict Correct 8.5 40.8

[1] "True: 69.4"

[1] "Confusion matrix of overall chance:"

Incorrect Correct

Predict Incorrect 0.0 0.0

Predict Correct 37.1 62.9

[1] "True: 62.9"

[1] "Confusion matrix of specific chance:"

Incorrect Correct

Predict Incorrect 26.1 10.7

Predict Correct 11.0 52.1

[1] "True: 78.2"

eval\_results\_SMEtag\_M\_20\_1\_fold\_1\_times\_first\_attempts.RData

> load(f)

> show.eval(eval.results)

Number of observations used: 29624

M = 20 eta = 0

-LL = 0.556 (overall and problem-specific learned chance would give 0.476 and 0.343 respectively)

-LL correct = 0.643 (overall and problem-specific learned chance would give 0.335 and 0.266 respectively)

-LL incorrect = 0.41 (overall and problem-specific learned chance would give 0.714 and 0.472 respectively)

MAE = 0.433 (overall and problem-specific learned chance would give 0.467 and 0.328 respectively)

RMSE = 0.52 (overall and problem-specific learned chance would give 0.483 and 0.405 respectively)

[1] "Confusion matrix:"

Incorrect Correct

Predict Incorrect 28.6 36.8

Predict Correct 8.6 26.0

[1] "True: 54.6"

[1] "Confusion matrix of overall chance:"

Incorrect Correct

Predict Incorrect 0.0 0.0

Predict Correct 37.2 62.8

[1] "True: 62.8"

[1] "Confusion matrix of specific chance:"

Incorrect Correct

Predict Incorrect 19.2 8.3

Predict Correct 17.9 54.6

[1] "True: 73.8"

eval\_results\_SMEtag\_M\_20\_1\_fold\_1\_times\_last\_attempts.RData

> load(f)

> show.eval(eval.results)

Number of observations used: 29624

M = 20 eta = 0

-LL = 0.537 (overall and problem-specific learned chance would give 0.404 and 0.297 respectively)

-LL correct = 0.573 (overall and problem-specific learned chance would give 0.206 and 0.169 respectively)

-LL incorrect = 0.426 (overall and problem-specific learned chance would give 1.006 and 0.686 respectively)

MAE = 0.421 (overall and problem-specific learned chance would give 0.373 and 0.275 respectively)

RMSE = 0.51 (overall and problem-specific learned chance would give 0.432 and 0.371 respectively)

[1] "Confusion matrix:"

Incorrect Correct

Predict Incorrect 17.9 36.0

Predict Correct 6.9 39.2

[1] "True: 57.1"

[1] "Confusion matrix of overall chance:"

Incorrect Correct

Predict Incorrect 0.0 0.0

Predict Correct 24.8 75.2

[1] "True: 75.2"

[1] "Confusion matrix of specific chance:"

Incorrect Correct

Predict Incorrect 6.1 2.1

Predict Correct 18.7 73.1

[1] "True: 79.2"

eval\_results\_SMEtag\_M\_20\_3\_fold\_10\_times\_all\_attempts.RData

> load(f)

> show.eval(eval.results)

Number of observations used: 561780

M = 20 eta = 0

-LL = 0.421 (overall and problem-specific learned chance would give 0.476 and 0.3 respectively)

-LL correct = 0.431 (overall and problem-specific learned chance would give 0.335 and 0.245 respectively)

-LL incorrect = 0.404 (overall and problem-specific learned chance would give 0.715 and 0.395 respectively)

MAE = 0.332 (overall and problem-specific learned chance would give 0.467 and 0.283 respectively)

RMSE = 0.444 (overall and problem-specific learned chance would give 0.483 and 0.377 respectively)

[1] "Confusion matrix:"

Incorrect Correct

Predict Incorrect 28.4 21.9

Predict Correct 8.7 40.9

[1] "True: 69.4"

[1] "Confusion matrix of overall chance:"

Incorrect Correct

Predict Incorrect 0.0 0.0

Predict Correct 37.1 62.9

[1] "True: 62.9"

[1] "Confusion matrix of specific chance:"

Incorrect Correct

Predict Incorrect 26.2 11.0

Predict Correct 10.9 51.9

[1] "True: 78.1"

eval\_results\_SMEtag\_M\_20\_3\_fold\_10\_times\_first\_attempts.RData

> load(f)

> show.eval(eval.results)

Number of observations used: 296240

M = 20 eta = 0

-LL = 0.56 (overall and problem-specific learned chance would give 0.476 and 0.345 respectively)

-LL correct = 0.642 (overall and problem-specific learned chance would give 0.335 and 0.267 respectively)

-LL incorrect = 0.42 (overall and problem-specific learned chance would give 0.714 and 0.478 respectively)

MAE = 0.435 (overall and problem-specific learned chance would give 0.467 and 0.328 respectively)

RMSE = 0.521 (overall and problem-specific learned chance would give 0.483 and 0.406 respectively)

[1] "Confusion matrix:"

Incorrect Correct

Predict Incorrect 27.9 35.8

Predict Correct 9.3 27.0

[1] "True: 54.9"

[1] "Confusion matrix of overall chance:"

Incorrect Correct

Predict Incorrect 0.0 0.0

Predict Correct 37.2 62.8

[1] "True: 62.8"

[1] "Confusion matrix of specific chance:"

Incorrect Correct

Predict Incorrect 19.5 8.9

Predict Correct 17.7 54.0

[1] "True: 73.5"

eval\_results\_SMEtag\_M\_20\_3\_fold\_10\_times\_last\_attempts.RData

> load(f)

> show.eval(eval.results)

Number of observations used: 296240

M = 20 eta = 0

-LL = 0.542 (overall and problem-specific learned chance would give 0.404 and 0.3 respectively)

-LL correct = 0.572 (overall and problem-specific learned chance would give 0.206 and 0.17 respectively)

-LL incorrect = 0.451 (overall and problem-specific learned chance would give 1.006 and 0.693 respectively)

MAE = 0.421 (overall and problem-specific learned chance would give 0.373 and 0.276 respectively)

RMSE = 0.512 (overall and problem-specific learned chance would give 0.432 and 0.372 respectively)

[1] "Confusion matrix:"

Incorrect Correct

Predict Incorrect 17.4 34.7

Predict Correct 7.4 40.5

[1] "True: 57.9"

[1] "Confusion matrix of overall chance:"

Incorrect Correct

Predict Incorrect 0.0 0.0

Predict Correct 24.8 75.2

[1] "True: 75.2"

[1] "Confusion matrix of specific chance:"

Incorrect Correct

Predict Incorrect 6.4 2.5

Predict Correct 18.4 72.7

[1] "True: 79.1"

eval\_results\_ONE\_KC\_autotag\_M\_20\_1\_fold\_1\_times\_all\_attempts.RData

> load(f)

> show.eval(eval.results)

Number of observations used: 172410

M = 20 eta = 0

-LL = 0.36 (overall and problem-specific learned chance would give 0.366 and 0.207 respectively)

-LL correct = 0.087 (overall and problem-specific learned chance would give 0.166 and 0.117 respectively)

-LL incorrect = 1.417 (overall and problem-specific learned chance would give 1.142 and 0.555 respectively)

MAE = 0.254 (overall and problem-specific learned chance would give 0.326 and 0.19 respectively)

RMSE = 0.401 (overall and problem-specific learned chance would give 0.404 and 0.308 respectively)

[1] "Confusion matrix:"

Incorrect Correct

Predict Incorrect 0.4 0.4

Predict Correct 20.1 79.1

[1] "True: 79.5"

[1] "Confusion matrix of overall chance:"

Incorrect Correct

Predict Incorrect 0.0 0.0

Predict Correct 20.5 79.5

[1] "True: 79.5"

[1] "Confusion matrix of specific chance:"

Incorrect Correct

Predict Incorrect 11.3 5.3

Predict Correct 9.2 74.2

[1] "True: 85.5"

eval\_results\_SCRAMBLED\_autotag\_M\_20\_1\_fold\_1\_times\_all\_attempts.RData

> load(f)

> show.eval(eval.results)

Number of observations used: 172410

M = 20 eta = 0

-LL = 0.422 (overall and problem-specific learned chance would give 0.366 and 0.207 respectively)

-LL correct = 0.358 (overall and problem-specific learned chance would give 0.166 and 0.117 respectively)

-LL incorrect = 0.67 (overall and problem-specific learned chance would give 1.142 and 0.555 respectively)

MAE = 0.336 (overall and problem-specific learned chance would give 0.326 and 0.19 respectively)

RMSE = 0.451 (overall and problem-specific learned chance would give 0.404 and 0.308 respectively)

[1] "Confusion matrix:"

Incorrect Correct

Predict Incorrect 9.7 20.5

Predict Correct 10.9 59.0

[1] "True: 68.7"

[1] "Confusion matrix of overall chance:"

Incorrect Correct

Predict Incorrect 0.0 0.0

Predict Correct 20.5 79.5

[1] "True: 79.5"

[1] "Confusion matrix of specific chance:"

Incorrect Correct

Predict Incorrect 11.3 5.3

Predict Correct 9.2 74.2

[1] "True: 85.5"