Andrew Hautau

Intro To Data Science

Class 7 Assignment

1. <https://raw.github.com/CrossTheStreams/data_science/master/class_7/sparse_matrix.r>
2. Performance metrics will over perform in training data, since models use training data to form a hypothesis. Only when these hypotheses can be evaluated against the test data can there be a real evaluation of the models.
3. Training data is used to inform models/hypotheses that can be evaluated against a test dataset. Test data must be independent of the training data and have the same probability distribution.
4. Test accuracy: 90/100 = 0.9 (90%)

|  |  |  |  |
| --- | --- | --- | --- |
|  | Condition Positive (Cold) | Condition Negative (Healthy) |  |
| Test Outcome Positive | True Positive  85 | False Positive  5 | Positive Predictive Value  (Precision)  85/(85 + 5) = 0.944  (94%) |
| Test Outcome Negative | False Negative  3 | True Negative  7 | Negative Predictive Value  7/(3+7) = 0.7  (70%) |
|  | Sensitivity  (Recall)  85/(85+3) = 0.9659  (97%) | Specificity  5/(5+7) = 0.4166  (42%) |  |