README

Changes since Milestone 4

- From-scratch and Pre-trained Models
    - Improved centering and scaling
    - Removed model checkpoints
    - Added TensorBoard logging
- From-scratch Model
    - Added glorot initialization to the from scratch model
- Pre-trained Model
    - Increased patience from 20 to 25
    - Increased maximum epochs from 100 to 250
    - Decreased number of units in output layers from 4,096 to 1024
    - Reduced learning rate from 0.1 to 0.001
    - Reduced starting momentum from 0.9 to 0.5

Changes since Milestone 3

- Both People and language Models
    - Added gradient boosting classifier to both datasets
    - Added precision, recall, and Hamming loss metrics to all models
    - Added multi-plot figure to visualize metrics
    - Added ‘musical’ to genres set for completeness
    - Added stacking to language and people models
- Language Model
    - Changed data source from TMDb to IMDb for consistency
    - Reduced number of principal components in QDA models from 100 to 10 to reduce overfitting
    - Changed count vectorizer to TF-IDF vectorizer for performance
    - Increased number of trees in random-forest models from 20 to 40 for performance
    - Reduced regularization parameter value in QDA models from 0.25 to 0 for performance
- People Model
    - Reduced number of principal components in QDA models from 100 to 20 to reduce overfitting
    - Increased number of trees in random-forest models from 20 to 60 for performance
    - Reduced regularization parameter value in QDA models from 0.25 to .075 for performance

Changes since Milestone 2

- Implemented multi-label approach to Y

Changes since Milestone 1

- Re-organized some data fields in pre-processed data