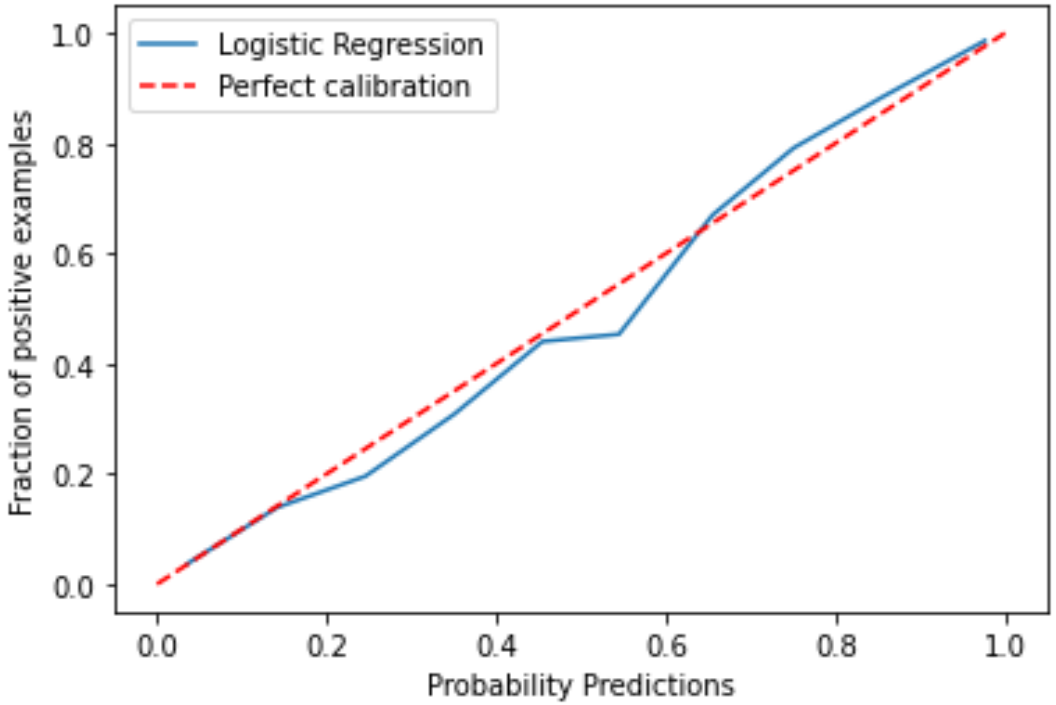
Train In Data

Probability Calibration Curves

Calibration Curve

- X axis : predicted probabilities
  - In bins or intervals

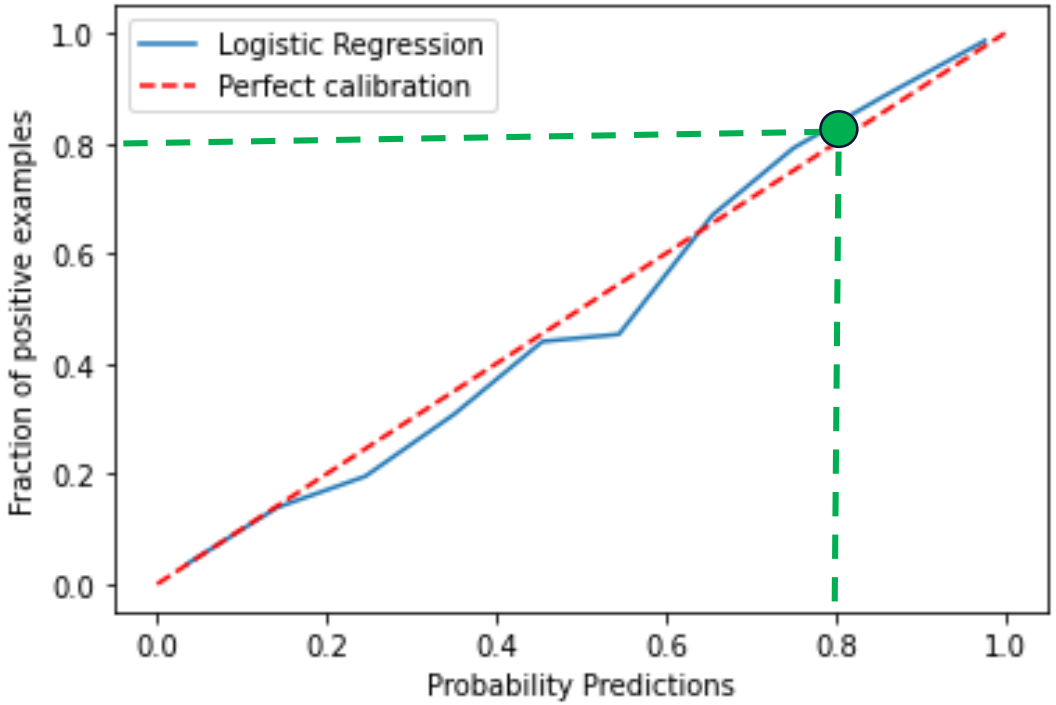
- Y axis : fraction of positive observations



Calibration Curve

- **X axis** : predicted probabilities
  - In bins or intervals

- **Y axis** : fraction of positive observations

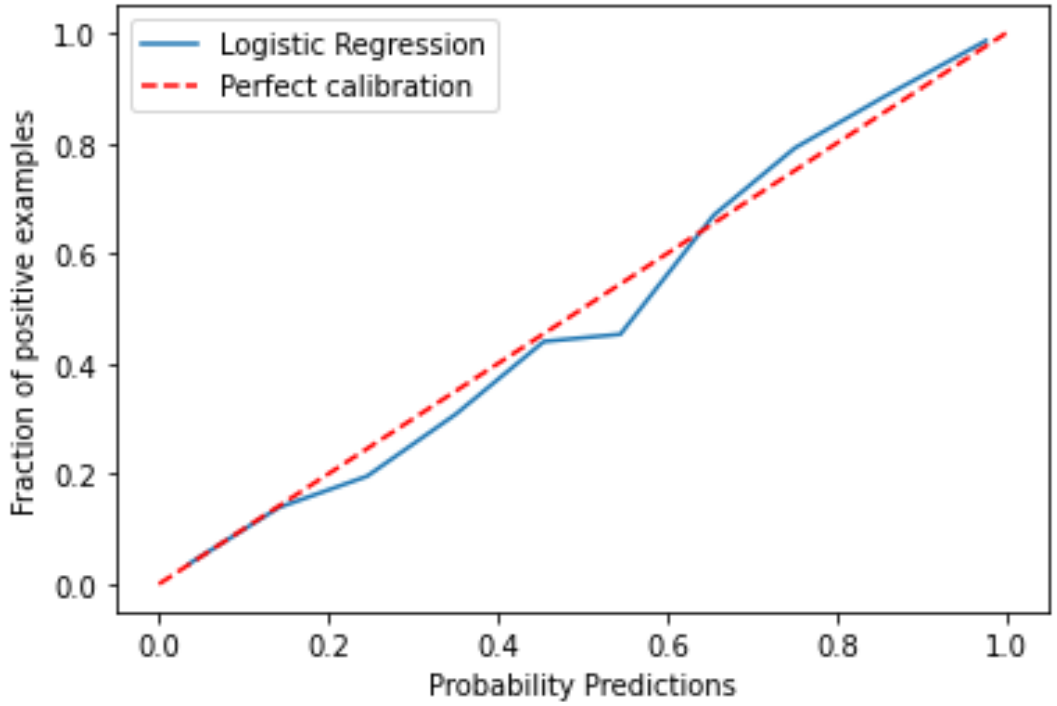


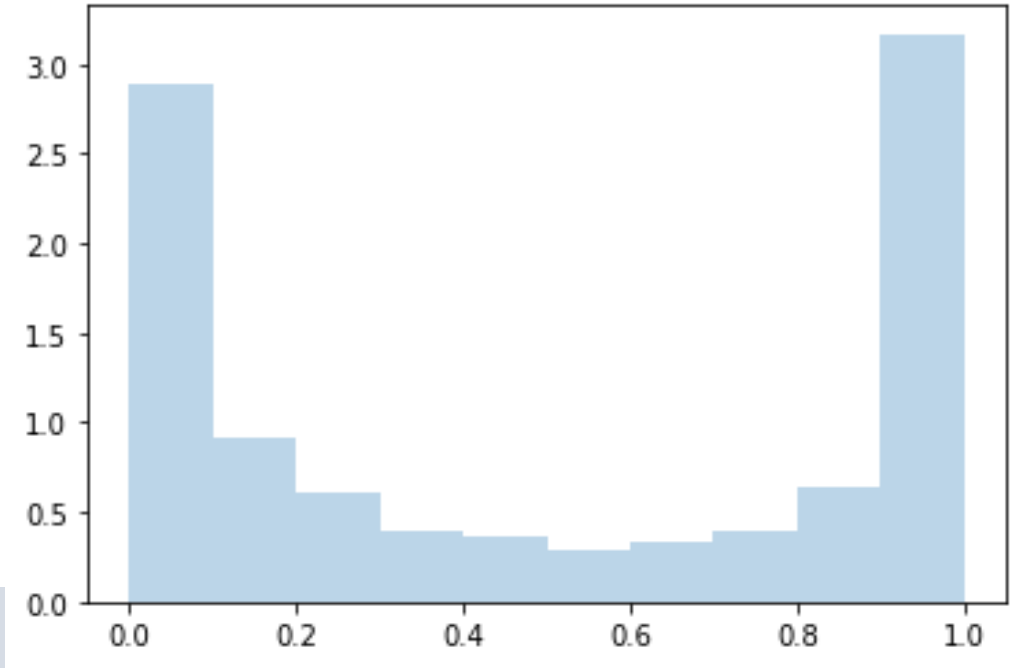
Calibration Curve

- The proximity to the red line is important

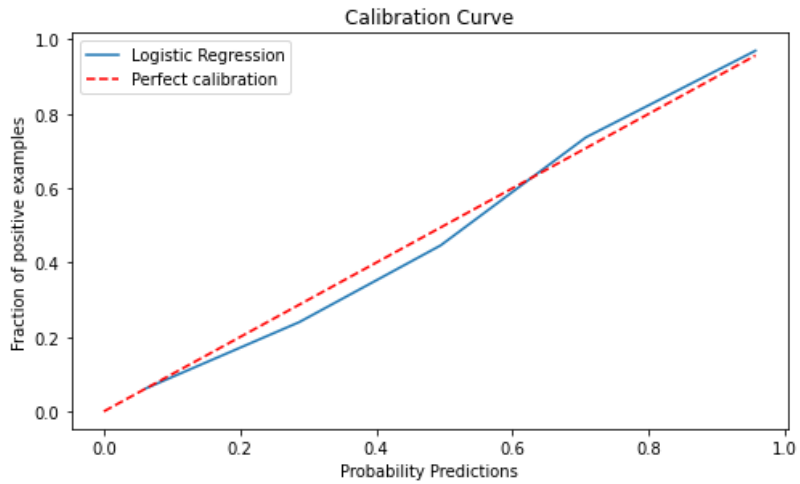
- The number of observations per interval is important

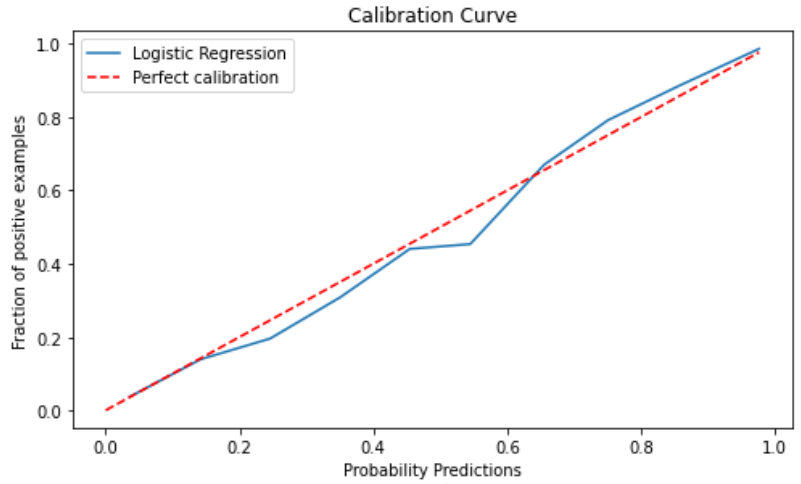
- Harder to obtain the real positive fraction if we have few observations

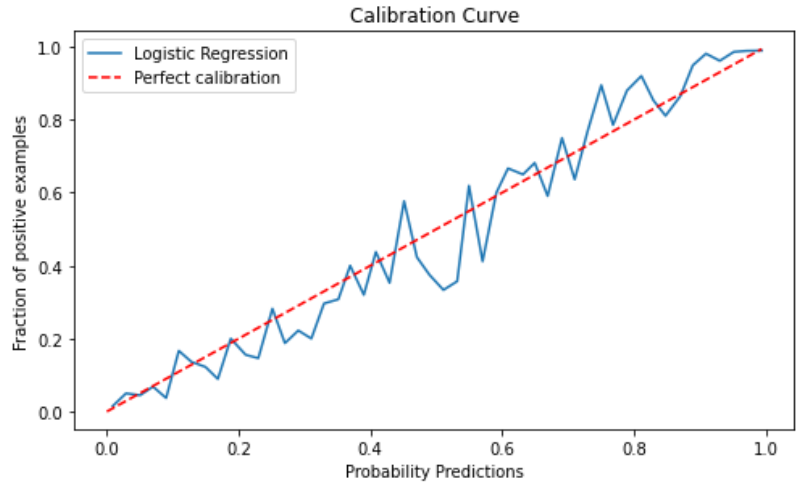


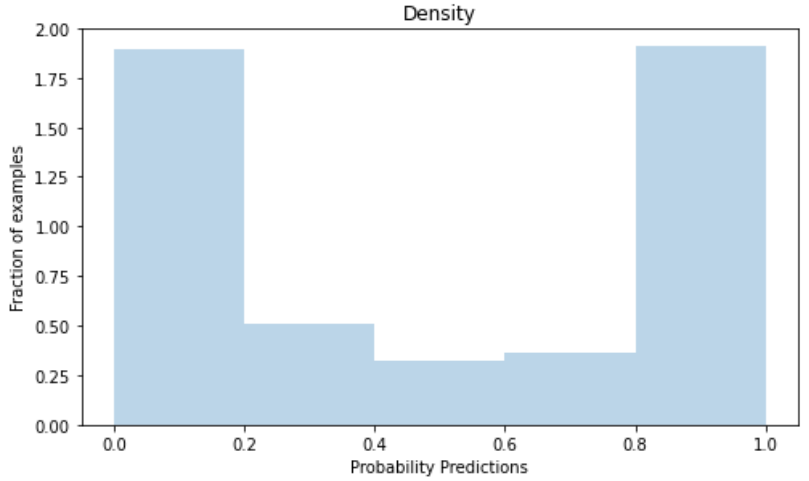


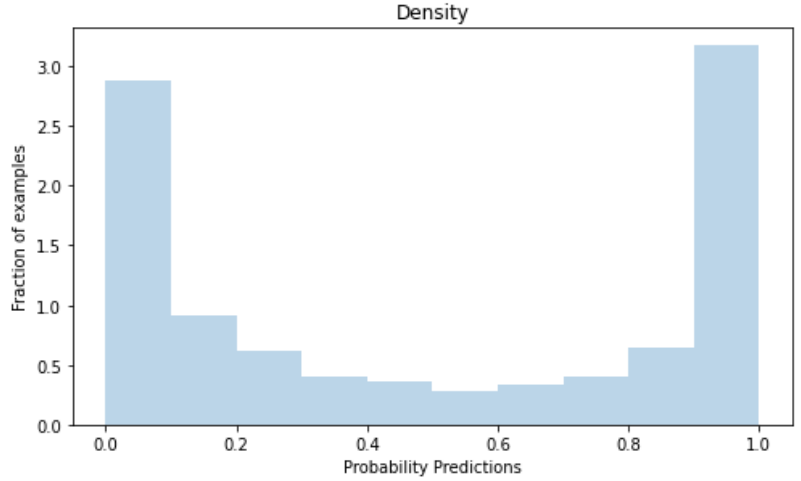
Calibration Curve – Balanced data

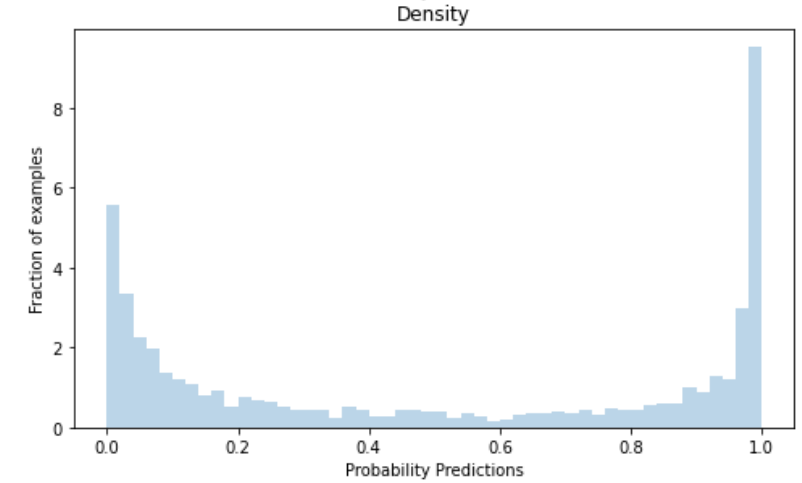








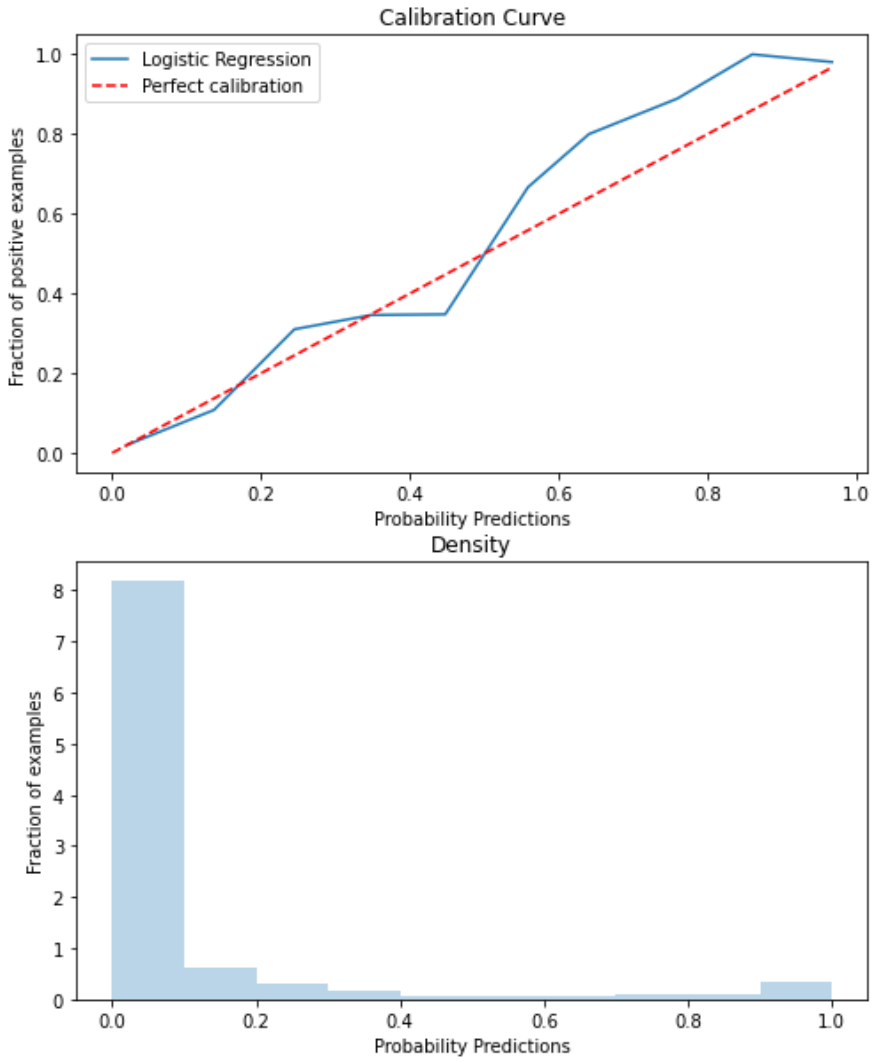


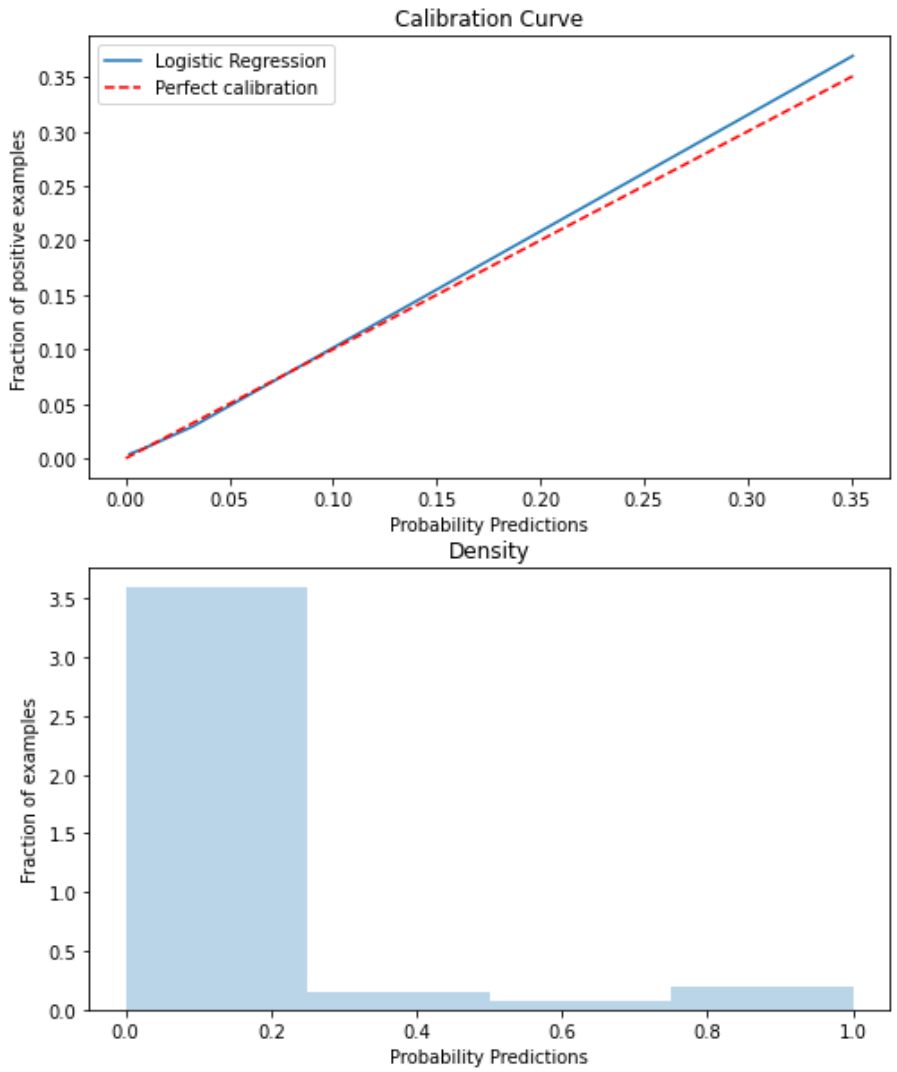


Calibration Curve – Balanced data

- Too few bins may look well fit
- Too many bins may look noisy
- We need the right balance of bins

Calibration Curve – Imbalanced Data





Calibration Curve – Imbalanced Data

With imbalanced datasets it is very hard to say if the probability is calibrated, because there are few observations of the positive class.

Increase the size of the dataset

Train In Data

THANK YOU

www.trainindata.com