Probability Calibration

Output of classifiers

- The class

- The probability of being of class 1

Probability and confidence

When performing classification we may want to predict the class label, and also to obtain a probability, certainty or confidence around the respective label.

Probability as certainty

- Probabilities can be much more informative than labels:

- "The model predicts this claim is fraudulent" vs "The model predicts this claim is 90% likely to be fraudulent"

- To convey likelihood, we need calibrated probabilities

Calibrated Probability: what is it?

- In calibrated probabilities, the probability reflects the true likelihood

- Calibration is the concordance of predicted probabilities with the occurrence of positive cases

- If 10 observations obtain a probability of 0.8 and the probability is calibrated, I expect around 8 of those to belong to the positive class

Calibration Curve

- If the probability is calibrated, we should see a match between the number of positive cases and the predicted probability

https://amueller.github.io/COMS4995-s20/slides/aml-10-calibration-imbalanced-data/#5

Calibration Curve

- If the probability is calibrated, we should see a match between the number of positive cases and the predicted probability

https://amueller.github.io/COMS4995-s20/slides/aml-10-calibration-imbalanced-data/#5
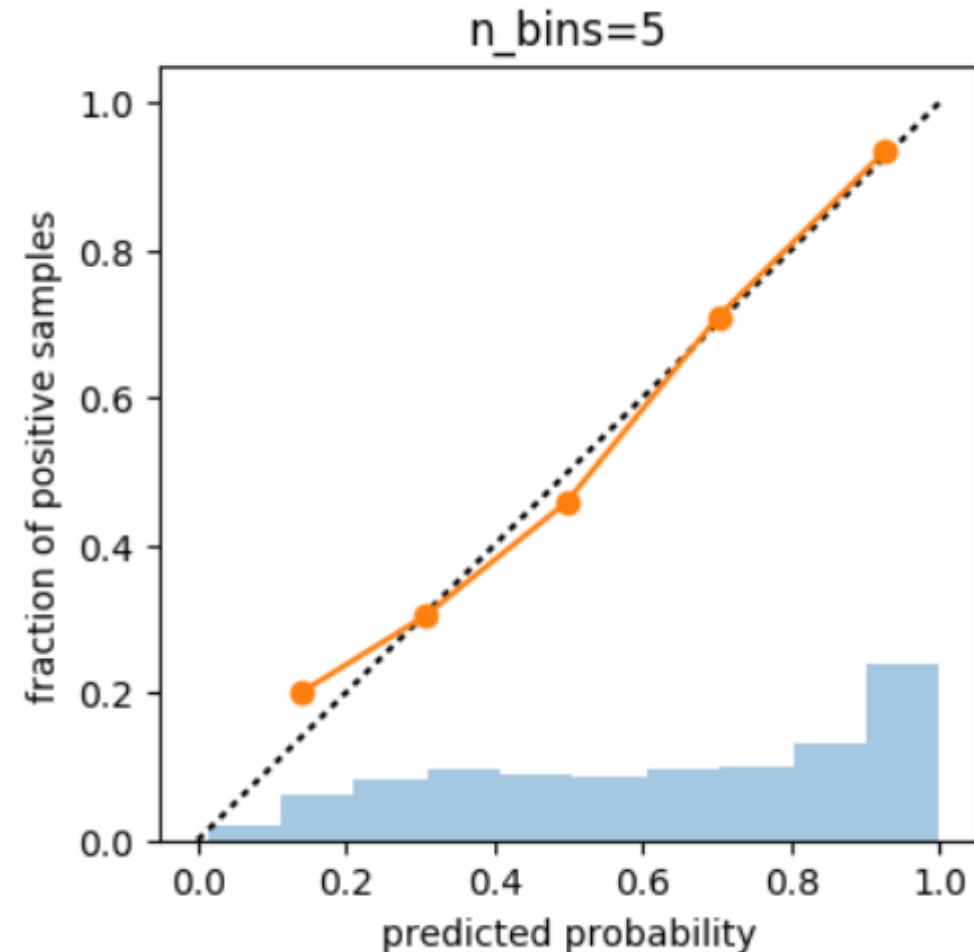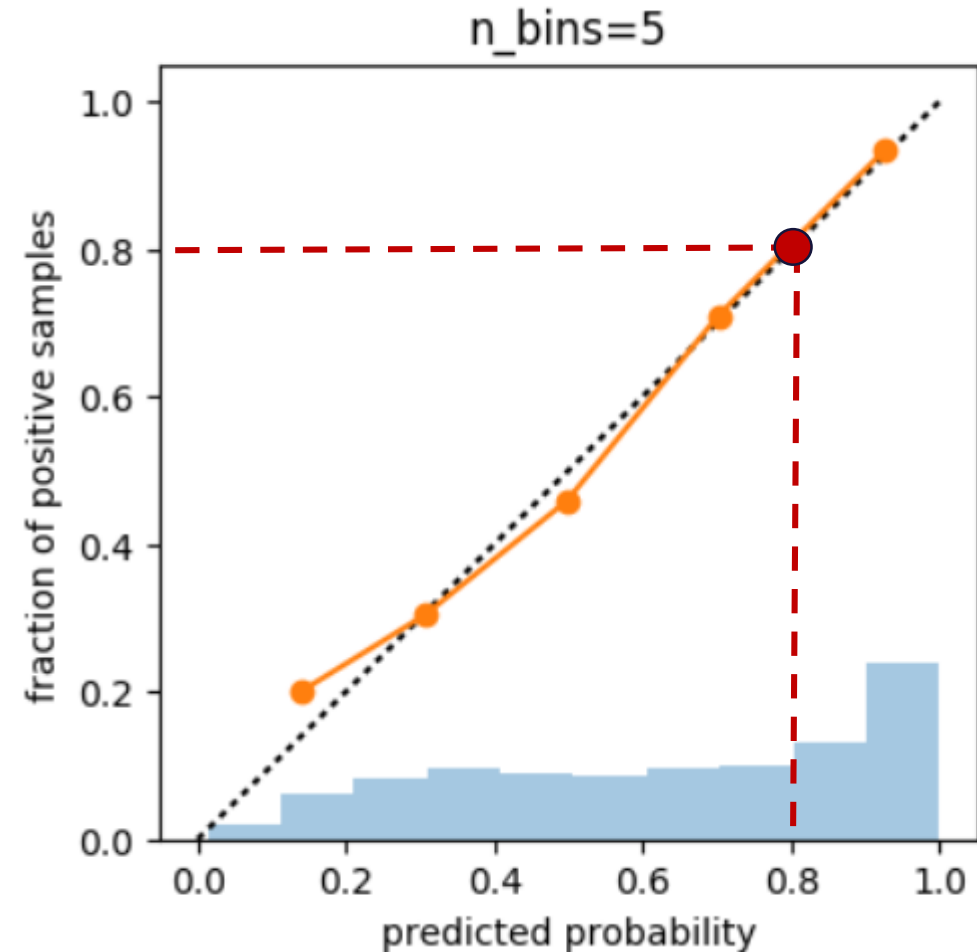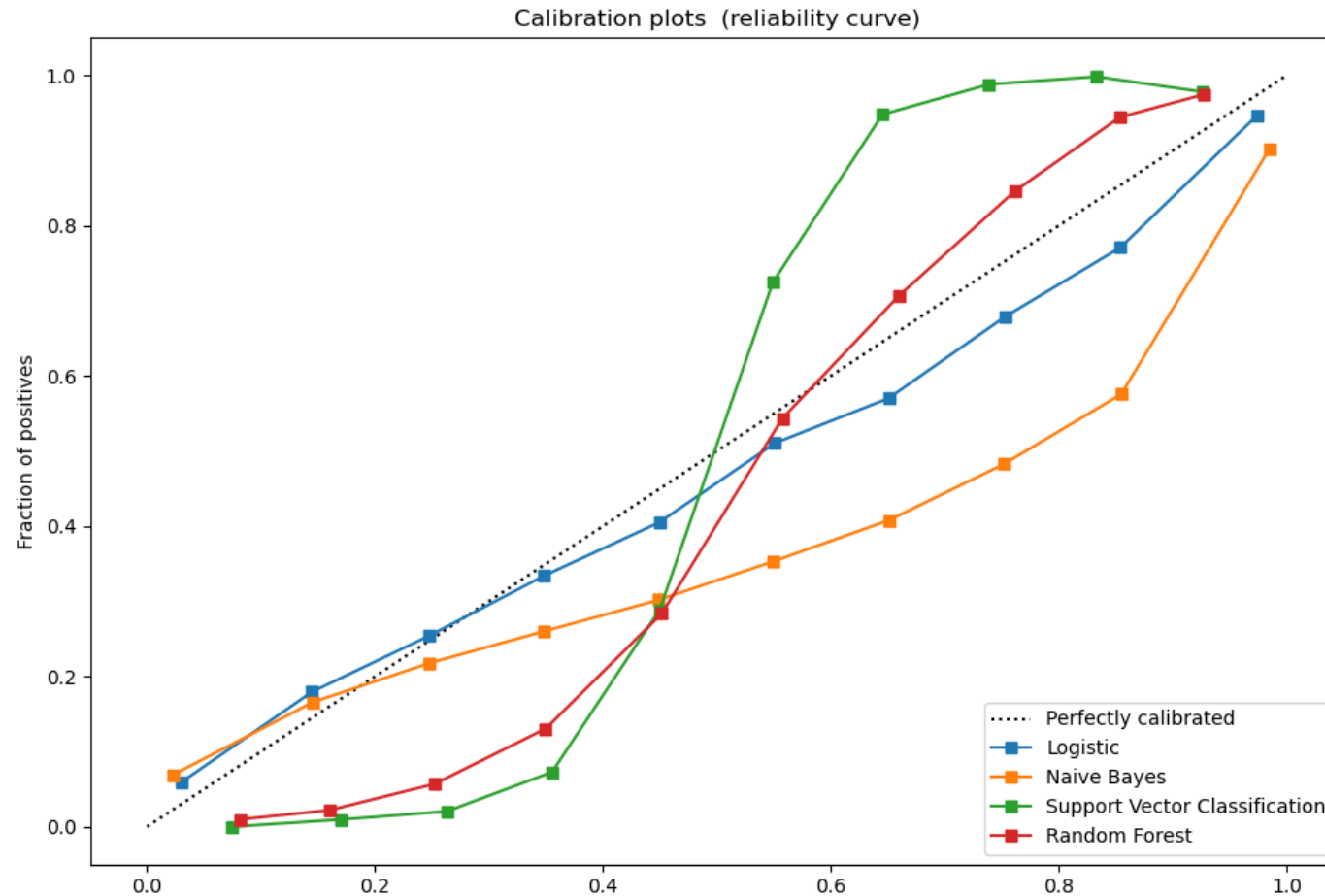
ML models and probability

- Logistic regression returns calibrated probabilities by design

- Some machine learning models return uncalibrated probabilities
  - ➢ Decision trees
  - ➢ Naïve Bayes

- Some models do not support probability prediction

ML models and probability

- Logistic regression returns calibrated probabilities by design

- Boosted algorithms, Random Forests and Naïve Bayes, push probability mass away from 0 and 1 yielding sigmoid shape distortion in the predicted probabilities

- Some models output the class with no probability, eg, SVMs

Train In Data

ML models and probability

Calibration plots (reliability curve)

https://scikit-learn.org/stable/modules/calibration.html

Uncalibrated probability: should we care?

To determine model performance, not necessarily.

As users of the model, we might.

✓ This patient is 80% likely to have cancer is more useful to a doctor than 0 or 1, the class prediction.

✓ This claim is 70% likely to be fraudulent is more useful to fraud investigators than 0.2, the uncalibrated prediction.

Train In Data