Probset #2

2. (b)

The probabilities outputted by a model match empirical observation

If you have a binary classification model that is perfectly calibrated - that is, the property we just proved holds for any $(a,b) \subset [0,1]$ - does this necessarily imply that the model achieves perfect accuracy? Is the converse necessarily true?

(i) perfectly calibrated $\rightarrow$ model achieves perfect accuracy

(ii) converse: model achieves perfect accuracy $\rightarrow$ perfectly calibrated

Statement: If model achieves perfect accuracy, model is not perfectly calibrated.

Proof:

If model is perfectly calibrated, model doesn't achieve perfect accuracy.

Contradiction: If model achieves perfect accuracy, then model is not perfectly calibrated.

$(a,b) \subset (0,1)$ so, if $(a,b) = (0.5,1)$:

└ model always predicts as positive

$$\frac{\sum_{i \in I_{a,b}} P(y^{(i)} = 1 \mid x^{(i)}; \theta)}{|\{i \in I_{a,b}\}|} < 1 \qquad \frac{\sum_{i \in I_{a,b}} \mathbb{I}\{y^{(i)} = 1\}}{|\{i \in I_{a,b}\}|} = 1$$

└ only positive labels exist

$\neq$

$\therefore$ model is not perfectly calibrated.
(proved that contradiction of (i) is false.)
(also proved that statement is true, and (ii) is false.)

Therefore, both (i) and (ii) are false.