Responsible Machine Learning

Exercise set #2

Exercise 1 - Consider we have binary classification data distributed as the pair (X, S, Y) where (X, S) is the feature vector and the distribution of Y given (X, S) is a Bernoulli distribution. Assume we have a risk score function f taking a finite number B of values in $[0, 1]$.

1. Provide the definition of calibration within group $S = s$ in each risk score value $b \in \{1, \ldots, B\}$ for the risk score function f in terms of conditional probabilities.

2. By suming over $\{1, \ldots, B\}$, write down the identity that relates $\mathbb{E}(f|S = t, Y = y)$ and $\mathbb{P}(Y = 1|S = t)$.

3. Under which conditions on the distribution of (X, S, Y), is this identity satisfied.

Exercise 2 - Consider binary classification data $((X, S), Y)$ and a binary predictor f. We denote by FPR the false positive rate and by FNR the false negative rate. Set the base rate $p = \mathbb{P}(Y = +1)$.

1. Find α such that the following relation holds :

$$\text{FPR} = \frac{p}{1 - p} \cdot \frac{1 - \alpha}{\alpha} \cdot (1 - \text{FNR}) \ .$$

2. Now considering that there are two groups with respect to the sensitive variable and that $\mathbb{P}(Y = 1 \mid S = 1) \neq \mathbb{P}(Y = 1 \mid S = 0)$, what can you conclude by applying the previous identity to each group ?