Data Literacy

Exercise Sheet #9

due on Monday, 10 January 2022, 10am sharp

1. **EXAMple:** Consider an algorithm that assigns class labels $R \in \{0; 1\}$ to individuals described by features X and sensitive attributes A. Let the true class-labels of each individual be Y. Define what it means for the algorithm to fulfill the fairness criterion of *independence* (mentioned in the lecture).

2. **Theory Question:** Consider again the situation described in task 1 above (for simplicity, assume that all variables can take finitely many discrete possible values $y_i \in \mathscr{Y}, x_j \in \mathscr{X}, a_k \in \mathscr{A}$). Remember that the fairness criterion of *separation* is fulfilled if R is conditionally independent of A given Y, i.e.
$$p(R \mid Y, A) = p(R \mid Y),$$
and that the criterion of *sufficiency* is fulfilled if Y is independent of A given R, i.e.
$$p(Y \mid R, A) = p(Y \mid R).$$
Assume that all events in the joint distribution $p(A, R, Y)$ have nonzero probability, and that A (sensitive attributes) and Y (true label) are not independent of each other. Show that in such a situation, no algorithm can simultaneously fulfill *sufficiency* and *separation*.

   [Hint: Remember the rules of probability:
$$p(S, W) = p(S \mid W) \cdot p(W) \qquad \text{product rule}$$
$$p(S) = \sum_{w_i \in \mathscr{W}} p(S, w_i) \qquad \text{sum rule}$$
$$p(S \mid W) = \frac{p(W \mid S) \cdot p(S)}{\sum_{s_i \in \mathscr{S}} p(s_i, W)} \qquad \text{Bayes' theorem.}$$
To complete the proof, start with the assumption that both sufficiency AND separation are fulfilled. Using the rules above, first show that this implies $p(A \mid R, Y) = p(A)$. Then show that this implies that A is independent of both R AND Y, in contradiction to our assumption.]

3. **Practical Question:** You can find this week's practical exercise in `Exercise_09.ipynb`