Instructions. Do this quiz in partnership with exactly one other student. Write both your names at the top of this page. Discuss the answer to the question with each other, and then write your joint answer below the question. Use the back of the page if necessary. It is fine if you overhear what other students say, because you still need to decide if they are right or wrong. You have seven minutes.

Question. The section on importance weighting in the lecture notes says that the weight for each labeled training example \((x, y)\) should be

\[
\frac{p(s = 1)}{p(s = 1|x)}.
\]

But the section on covariate shift says that the weight should be

\[
\frac{p(s = 0|x)}{p(s = 1|x)}.
\]

Explain intuitively the reason for the different numerators.

Answer. The goal of importance weighting is to make labeled examples (those with \(s = 1\)) be representative of all examples, but the goal under covariate shift is to make labeled examples be representative only of unlabeled examples. In both cases, an example for which \(s = 0\) is more likely should have higher weight, which is achieved by the denominator. However, in the second case the increase in weight as \(p(s = 0|x)\) increases should be more extreme. The numerator \(p(s = 0|x)\) achieves this in comparison to the constant numerator \(p(s = 1)\).

Additional note. Different real-world scenarios have different availability of labeled data, and different learning goals. It is important to understand each real-world scenario carefully, and not to apply a method appropriate for one scenario to a different scenario.