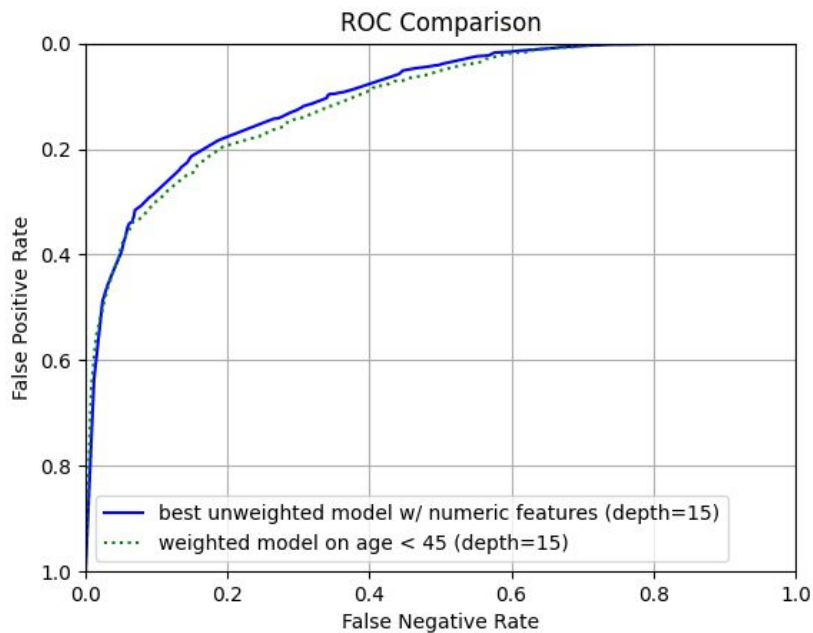


An ROC curve comparing a decision tree using the best hyperparameters you found in the previous assignment with a weighted decision tree using the same hyperparameters, but a sample weight of: 10 when the age is  $< 45$  ( $x[0]$  is the age feature) and 1 when age is  $\geq 45$ . Train both on the training set (no cross validation) and evaluate them on the validation data.



In 3-4 sentences, describe a condition when you might like to use this sample weighting with the Adult data set.

I would probably want to give higher weights if I wanted to focus the learning to a specific area (e.g., people under 45 years of age) and get the model to learn that with more focus. This is because the lower weighted samples would not contribute as much to the probability weights and so segmenting out such a sample would also not shift entropy by as much as a higher weighted sample.