

1 **Example of gain calculation in the XGBoost model**

2 An example of gain calculation for effect of GFR on AUC is provided belows; the goal being
3 to find a split that reduces the loss (mean squared error) the most by evaluating potential
4 splits on the GFR variable.

5 Suppose XGBoost considers splitting the data into two groups based on a GFR threshold of
6 60 mL/min/1.73m² (reduced renal function (<60) and normal or elevated renal function (≥60)):
7 calculate the total loss (sum of squared differences between the true AUC and the average
8 AUC for all patients) before the split and within each subgroup. The reduction in loss due to
9 the split (the gain) is the difference between the total loss before the split and the sum of the
10 losses for each subgroup after the split.

11 This process is repeated, evaluating multiple splits and using the concept of gain to build a
12 model that accurately predicts drug exposure based on renal function and the other
13 covariates. To ensure the model doesn't become overly complex (which could lead to
14 overfitting), regularization might penalize splits that add little predictive value relative to their
15 complexity, thus balancing accuracy with model simplicity.