

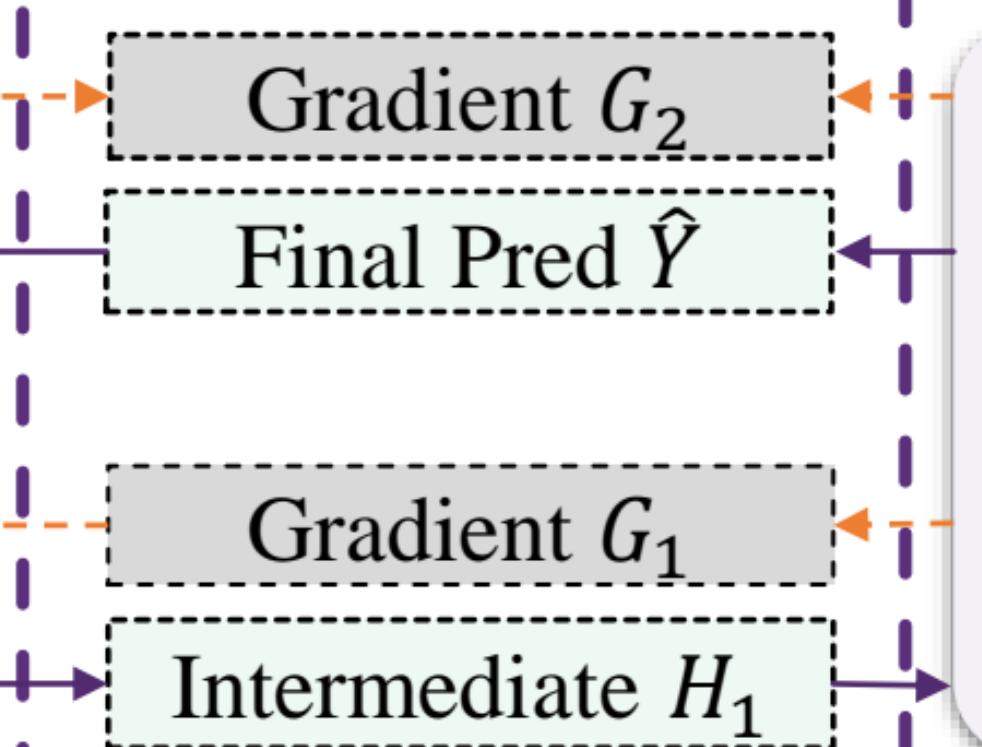
Data Party

Forward
Backward

Data X

Label $Y \rightarrow$ Loss L

Model Head M_{head}
embedding + n_{head} encoders



Model Party

Model Tail M_{tail}
 n_{tail} encoders + head layer