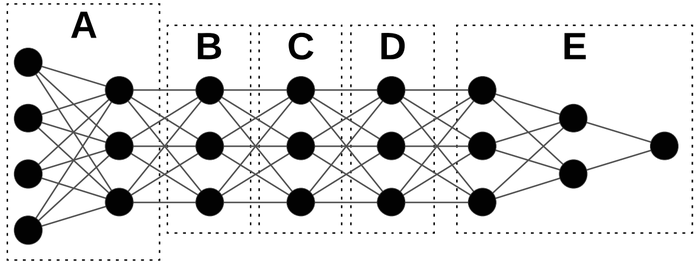
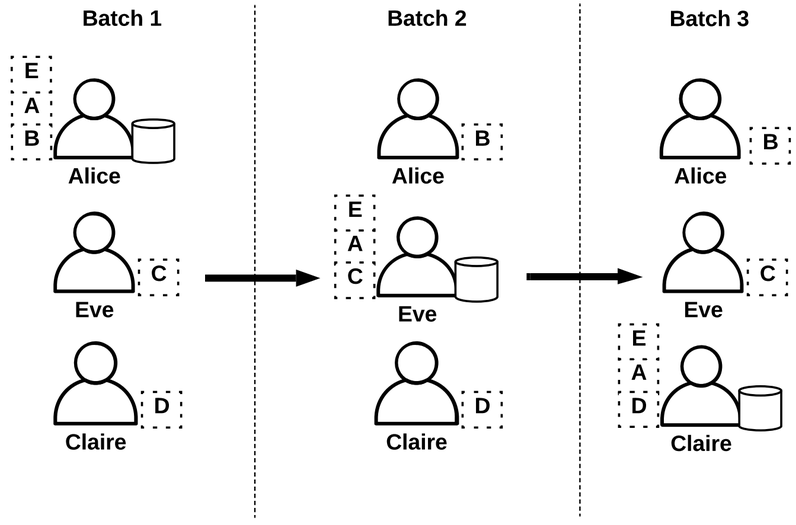
Split Neural Network





Vulnerability B

Eve can approximate the X values of Claire;

*- Eve chooses **X^** values which approximately span the domain of **X***

*- Eve recieves the activation signals produced by her **X** values*

*- Eve can statistcally correlate the values of A that she received with the **X^** values that she entered. This correlation could perhaps be established using a second learning model with **A** as its inputs and **X^** as its label.*

*- When it's Claires turn to train, Eve will receive **A** values from Claire.*

*- Using the model trained in the step before, Eve can make predictions as to the **X** values of Claire.*

Key Takeaway;

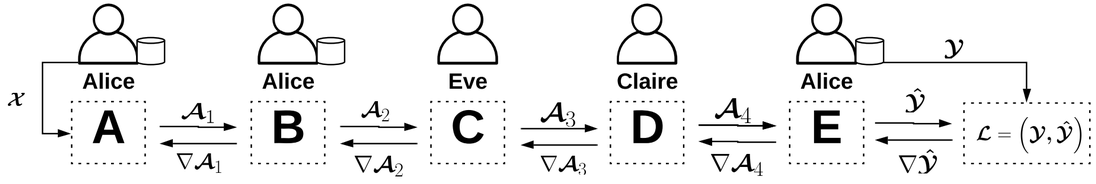
Data owners should not also be involved in processing the model as this can leak information.

If Eve is the Researcher, Vulnerability A applies.

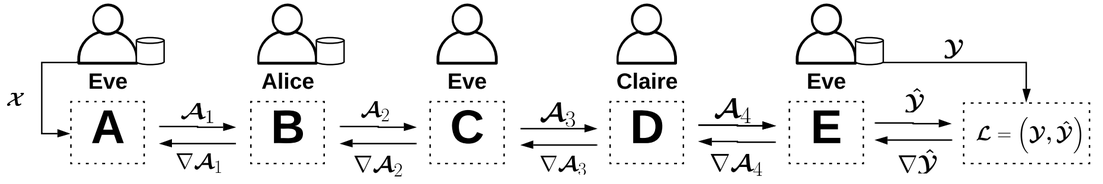
*- Just replace **M_0** with **B(A(X))***

+

1.



2.



3.