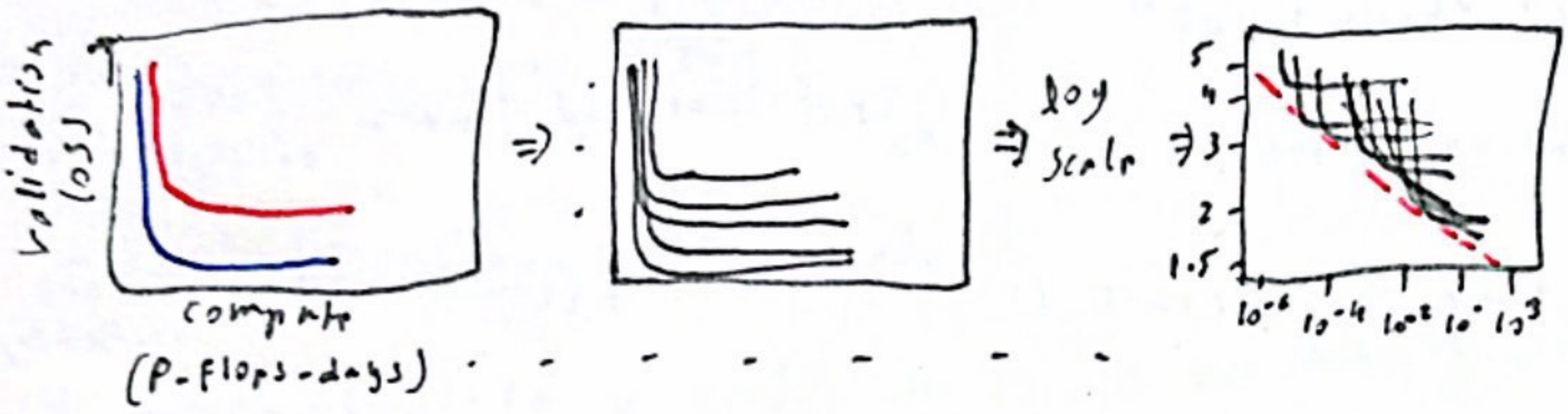
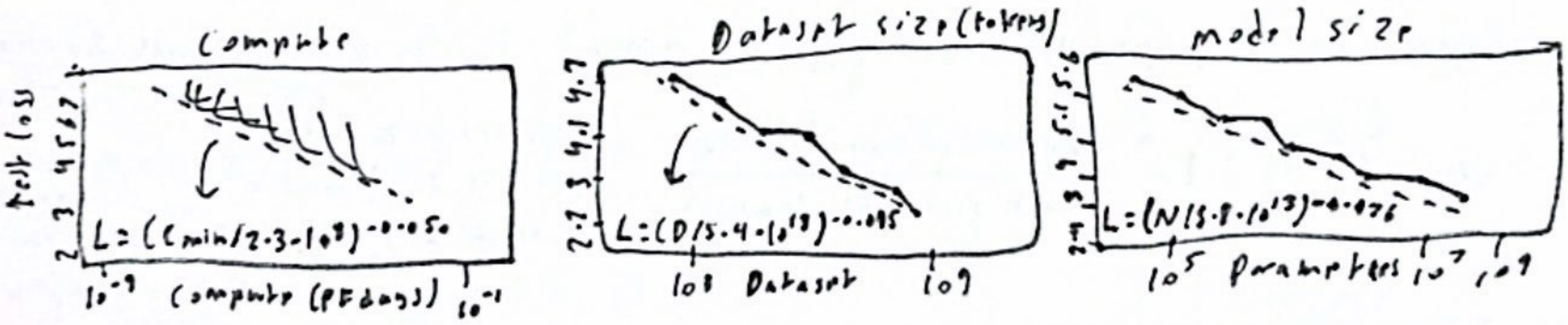
Nural scaling Laws

- as we train a AI model its error rate generally drops off Quickly and then levels off if we train a larger model it will achive a lower error rate but requires more compute scalling to larger and larger models we end up with a family of curves swiching to log scale a trend emerges. where no model can cross this line AKA the compute Efficency frontier



- This trend is one of three nural scalling laws that have been observed error rate scales in a simmilar way with compute, Model size and Data set size. and dose not depend much on model architecture or other Agorithmic details.



- on log plots these show straight lines and the slope of each line is equal to the exponent of fit equation (L) Larger exponents make for steeper lines and more rapid preformance improvement.

Feature importance.

mostly used in desicion trees

- take for ex desision trees they often overfit when unconstrained, meaning when we allow the tree to grow without any ristrictions (Like max depth, min samples per leaf etc) it can create a model thats too complex and overfits the training data, this leads to poor generalization to never unseen data basically memorizing the train set without learning the patterns. to solve this we can use. Feature importance is a way to mesure relitive importance of each feature in a desicion tree higher the. Feature importance the more important the feature for predictions we can use feature importance to identify important features used it works by looking at how much each feature contributes to reducing uncertainty in the data at each split in the tree.