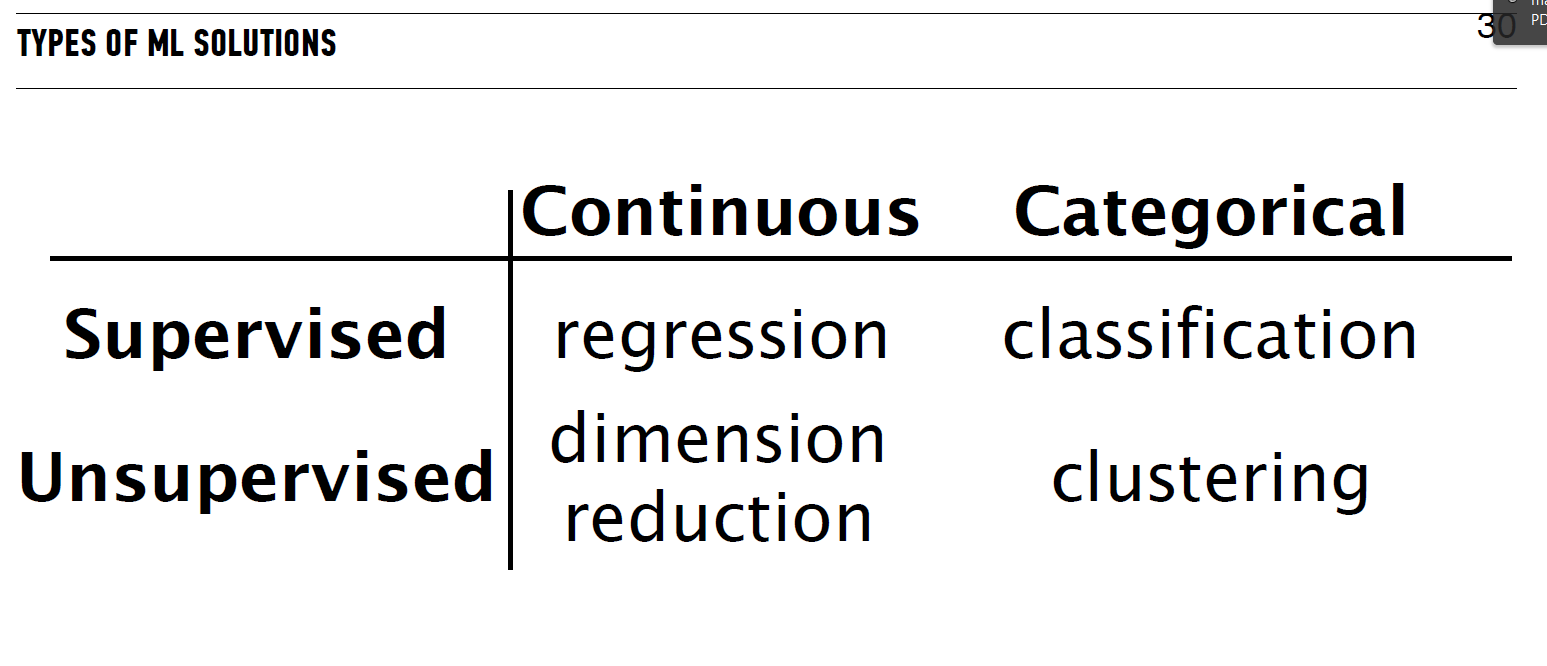
Data Science Lecture 4

Domains of ML

* Supervised – give solution for a subset of the data
* Unsupervised –

From Slides:



Overfitting – pushing the training error rate to zero.

Cross validation – taking the same ratio of testing and training and just switching it around. We have to make sure that we don’t reiterate with the same records. That is, we split the dataset into N distinct groups and test with each.