Use this as a guide to the lecture presentation and take notes on your own.

| Slide | Notes |
| --- | --- |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  | Schema/space – same thing |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  | A model is just a representation of the data, it is not the data. We need a model to make use of the data. Any model is not going to be representative of all the data. |
|  |  |
|  | The quality of the model is determined on how good it does versus the test set, the training set is irrelevant. |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  | I see a chair: we know there is a chair, there is an “I” and there is something that is connecting the two, the seeing. |
|  | Unstructured could also just be data that is not suitable to you. |
|  |  |
|  | Tweets are highly structured, 1-140 characters long, there is a from, you have a lot of intrinsic info with the source just being a “tweet” |
|  | ` |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |