Week1!
What: Machine Learning!
-A computer learns, when for some task T and performance measure P, the value P increases for T with experience E.(T. Mitchel)
 Note: does not explain how to learn and does not split data in train/test
-Learning = Representation + Evaluation + Optimization (P. Domingos)
 Note: does not include data

(representation), evaluation on train/test time is different.

10 Components of a Learning System

World->Data->Features: Representing textual and visual data!

-Data point i is represented by a vector xi !

-Machine learning is concerned with distances!

-Distances depend on the task / features!

-Example Nearest Neighbors!

Features->Training-Test: The goal of ML is to generalize
beyond the examples in the training set. Never evaluate the
final performance of the Machine Learning System on examples from the training set!

Getting is labels is expensive: expert annotations, laymen annotation, amazon mechanical turk,search engines, knowledges sources.!

Evaluation of Learning Model

Model on trainset was perfect, but evaluation on testset was bad. Diagnose the error: overfitting or wrong model?!

Bias: assumptions made by the learning model!
Variance: learning quirks from the data!

How to check for bias or variance

Train 15 times on subset->!
-Huge difference between run? -> High Variance -> Model unstable!
-Performance way off what I expect->Learning model too simple   -

>Representation not strong enough!
Low bias: Complexity of the model correct.!
Low variances: every time we run the program we get similar performance.

Reduce bias and variance
-Dimensionality reduction and feature selection

can decrease variance by simplifying the model. Similarly, a larger training set tends to decrease the variance.!
-Adding features tends to decrease bias, at the expense of introducing additional variance
Improving your ML system
-More training data !
-Review learning model (model or objective function