## Overarching Themes in this Area

## Recent Successes (last 3 years)

## Major Obstacles Impeding More Rapid Progress

## Areas of Neglect

## Strategic Priorities & Investments That Will Advance Innovation

## RAW NOTES BELOW

LargeScale Inference & Learning Breakout Session 4/20/16

SESSION 1:

What is large scale inference & learning?

— lots of data

— distributed learning

— large = scientist has trouble doing it

    — don't have the resources to do the inference for a certain amount of data

— given the resources you have, can you do it?

— runtime, not space?

— geoscience questions, and environmental research

    — NASA has started to create a platform for sharing data and resources

— this will impede progress if not implemented

— big data, small labels

    — even if you design something for big data, you might not have a large input or a large

number of labels.

— data is largely unlabeled

— academic datasets vary, nobody its large enough datasets to show that their algorithms are working. those datasets would be easier to get from an industrial context

— industry either does not want to or cannot release their datasets because of privacy concerns

— funding issue in academics. cannot fund largescale repositories that can be used for machine learning

— if you want an exabyte of data for research in academics, where do you go? nowhere

— HBC, if you want to do research on huge amounts of data, there are centers for that. not for Machine learning.

— only small amounts of data are going on the amazon cloud

— should push for more and more to be placed there

— Berkeley spent months perfecting computation so that the computing wasn't the bottleneck

— what is the point of wanting to use huge data for tests if you can test on smaller datasets and show that it will work on bigger datasets?

— cancer research needs large datasets

— genomic data is giant. you need to be able to do machine learning on huge datasets to do genomics