

Data Structures for Fast Statistics

An ICML 2004 Tutorial

Alexander Gray and Andrew Moore

School of Computer Science
Carnegie Mellon University

“This tutorial is proud to
offer **absolutely no**
statistical innovations”

- Alex Gray and Andrew Moore

FAQ

Q: *Why fast statistics/learning computations?*

- A: 1. Massive datasets → become feasible.
2. Faster ‘inner loop’ → allows more search
(iterations, restarts), cross-validation,
feature selection, bagging/boosting/stacking,
etc. → better learning.

Q: *Why data structures?*

- A: It's the data structures which allow certain
powerful classes of fast algorithms

FAQ

Q: *What are we going to cover?*

- A: We'll cover a selected set of common *core statistics/learning operations*. Many other statistics/learning methods are *analogous* to these operations OR have them as *subproblems*.

Q: *Where do we get more?*

- A: From 1) the references, 2) asking us.