# 数理工学コース　数理システム論分野　下平研
# （理化学研究所と兼任）

革新知能統合研究センター
Center for Advanced Intelligence Project

## 統計学，機械学習，データサイエンスの手法と理論を探求

統計学が注目されています。ビッグデータ、データマイニング、人工知能の流行を支える理論的基盤として統計学は重要な役割を果たしています。ランダムネスを考慮してデータから帰納的推論を行う方法論を提供することが統計学の大きな特徴です。ベイズ統計学の事後確率、頻度論の$p$-値など、不確実性のもとで信頼度を定量化する試みは科学・工学・医学など様々な分野に普及しました。確率モデルを通してデータから推測、予測、決定を行うための様々な手法や概念、たとえば最尤法、モデル選択、ロバスト統計学、漸近理論、ブートストラップ、仮説検定などが生み出されてきました。一方で、ウェブやソーシャルメディア、または生命科学や宇宙科学では大量のデータが主導する新しい方法論の必要性が増しています。

## 現実のデータにとりくんで，新たな理論を作る

かつて遺伝学においてR. A. Fisherが統計学を飛躍的に発展させたように、現実と向き合うことが方法論の発展をもたらします。研究室では、これまでにDNA配列解析、遺伝子発現解析でよく使われる統計手法を提案したり、機械学習の汎化誤差の理論、最近では因果推測，ネットワーク成長メカニズムの統計推測や、新しい情報統合の多変量解析法を提案してソーシャルメディアからの画像認識、文書データからの自然言語処理などの分野でも成果があります。

## 数学とプログラミング，どちらも重要

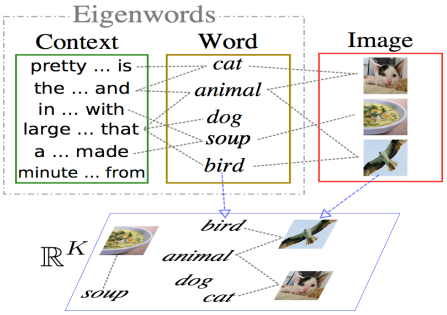
研究で最も重要なのはアイデアとデータです。そして数学とプログラミングは力です。定理の証明とコーディングは似た作業ですね。数学に自信のある人、Python，R，C++のスキルがある人は活躍するチャンスがあるし、やる気さえあれば研究を通して実力はつくものです。

## たとえば，こんな研究やってます

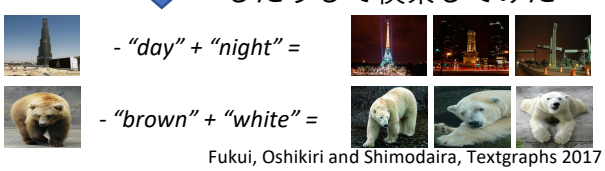
### グラフ（ネットワーク）を低次元に埋め込むデータ解析手法を考える

Shimodaira, Neural Networks 2016

↓ ソーシャルメディアのデータを埋め込んでみる

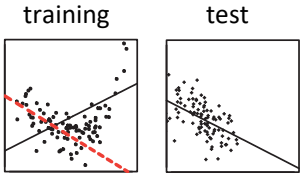



↓ 画像と単語を引いたり足したりして検索してみた




Fukui, Oshikiri and Shimodaira, Textgraphs 2017

### 学習とテストでデータの分布が変わるときの統計理論をひっそりと考える




確率密度比(density ratio)

$$w(x) = \frac{f_{\text{test}}(x)}{f_{\text{train}}(x)}$$

Shimodaira, Journal of Statistical Inference and Planning 2000

↓ 共変量シフトと命名！(covariate shift)

機械学習の分野でよく使われるようになる

Shimodaira (2000)の論文被引用数は700くらい

↓ しらないうちに．．．

ディープラーニングを加速する手法（Batch Normalization)に組み込まれる

Batch normalization: Accelerating deep network training by reducing internal covariate shift (Ioffe and Szegedy, ICML 2015)

彼らの論文被引用数はたった２年で４１００くらい．．．

http://stat.sys.i.kyoto-u.ac.jp