

### 統計学と機械学習の方法論を探求

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。

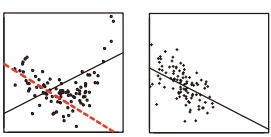
### 現実のデータにとりくんで、新たな理論を作る

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。

共変量シフト：予測分布と確率密度比




$$E_0 \left\{ \frac{q_1(x)}{q_0(x)} \log p(y|x; \theta) \right\} = E_1 \{ p(y|x; \theta) \}$$

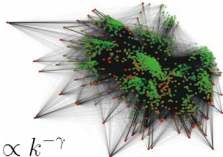
リサンプリング



GPGPU  
並列計算

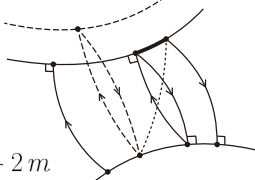


$P(k) \propto k^{-\gamma}$



複雑ネットワークの統計解析

情報幾何学

$$D(q, p) = - \int q(x) \log \frac{p(x)}{q(x)} dx$$


モデル選択

AIC = -2 log L(θ) + 2m

情報量規準・汎化誤差

機械学習 ディープラーニング

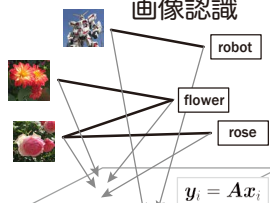
# 統計学

ベイズ統計学と頻度論をつなぐ  
仮説検定の高次漸近理論

$$P(\sigma^2) = \Phi \left\{ \sigma \Phi^{-1}(Q(\sigma^2)) \right\}$$

分子進化系統樹と遺伝子発現解析

画像認識




グラフの埋め込み

$$\phi(A) = \sum_{i=1}^N \sum_{j=1}^N w_{ij} \|Ax_i - Ax_j\|^2$$

情報統合の多変量解析

単語のベクトル表現



国名 (English) — 首都名 (English)  
国名 (Spanish) — 首都名 (Spanish)

自然言語処理