

アイスクリーム × データサイエンス で未来予測!?

立命館大学
理工学部数理科学科の
データサイエンスコースでは、
機械学習をはじめとする
データサイエンスの理論を
数学に基づいて学べます。

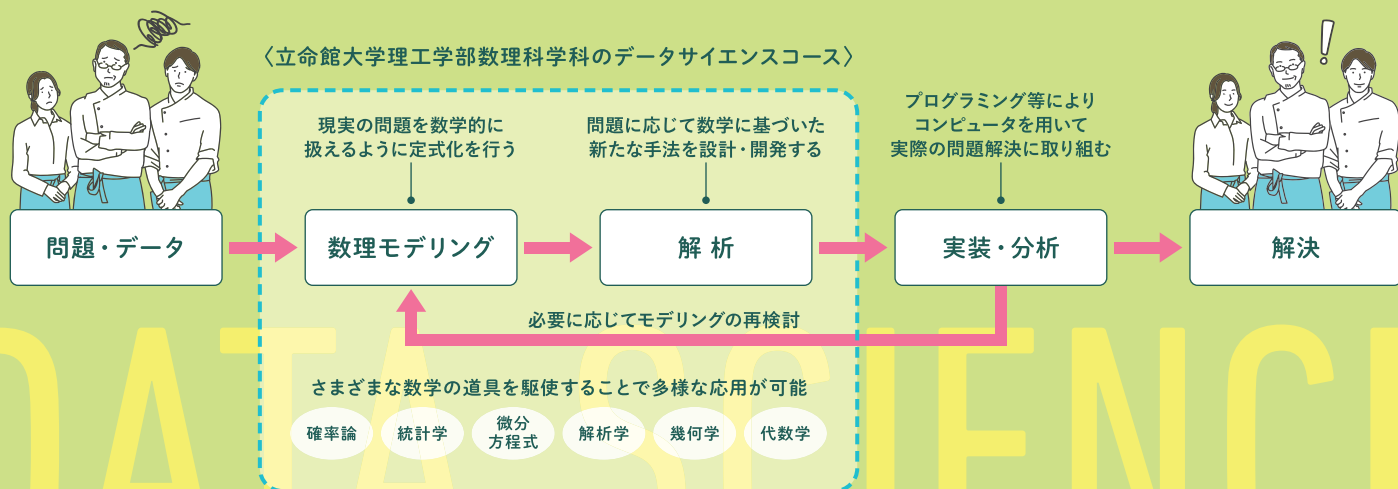
データサイエンスの
根幹を支える
さまざまな理論数学と
その応用への
習熟を目指して
教育を行っています!

立命館大学 理工学部 数理科学科

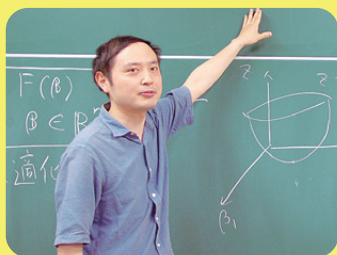
データサイエンスコース

さまざまな分野の数学が支える確かなデータサイエンス

データサイエンスで現実の問題を解決するためには、問題を数式を用いて表現し（数理モデリング）、その問題を解くための方法を数学を用いて創出し（解析）、コンピュータを用いて解き実際に役立てる（実装）というプロセスが重要です。立命館大学数理科学科のデータサイエンスコースでは特に数学を重視しており、数理モデリングや解析の部分を中心に学びます。プログラミングや応用についても扱いますが、それらも盤石な数学的基礎とともに進めることを目指します。データサイエンスでは多種多様な問題があり、確率論・統計学・微分方程式・解析学・幾何学・代数学など、いろいろな数学を駆使することで確実に幅広い応用が可能です。そのようなさまざまな分野の数学を学べるのが立命館のデータサイエンスコースです。



数理科学科 佐藤先生の解説



数理科学科 佐藤 寛之 准教授



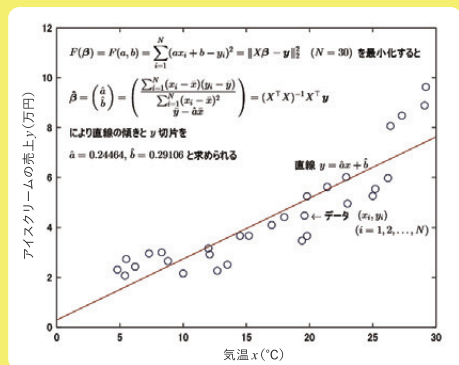
あるアイスクリーム店が、最高気温25℃と予報されている日にアイスクリームをどれだけ作るべきか検討しているとしましよう。そこで、過去のある30日分の気温と売り上げのデータから、25℃の日の売り上げを予測することにしました。

【図1】には、気温と売り上げのデータに加えて、気温 x ℃から売り上げ y 万円を予測するための直線を描いています。実は、この直線はグラフを見てそれらしい直線を引いたわけではなく、数学的に求めた妥当なものです。この直線の方程式 $y = \hat{a}x + \hat{b}$ の傾き $\hat{a}$ と y 切片 $\hat{b}$ を求めることは、ある2変数の2次関数 $F(a, b)$ の値が最小となる a と b を求める問題になります（数理モデリング）。関数 $F(a, b)$ のグラフは3次元空間では【図2】のように、またその値を高さと見なして2次元平面上で地形図の等高線に相当する曲線を描くと【図3】のようになります。【図1】中の $F(a, b)$ と $\hat{a}$ と $\hat{b}$ を表すための和の記号 Σ による式は、大学で学ぶ線形代数や微分積分学の知識を用いるとすっきりとした形で書かれ、また見通しよく求められます。こうして $\hat{a}$ と $\hat{b}$ が求まると、25℃の日の売り上げは $(25\hat{a} + \hat{b})$ 万円くらいだろうと予測できます。このような手法を回帰分析といいます。

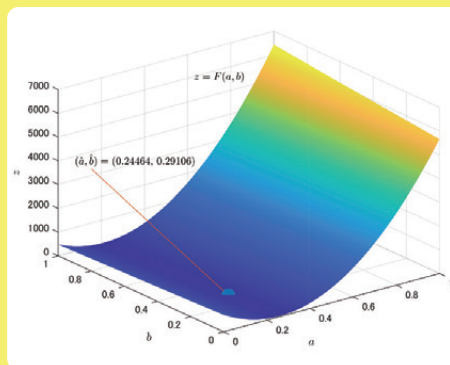
ただ、2変数の2次関数の最小を調べればどんな予測も可能となるというわけではありません。このアイスクリームの例にしても、実際には気温だけでなく湿度も売り上げに関係するでしょう。他にもいろいろな指標を考慮すると、とても多くの変数をもつ関数の最小について調べなくてはなりません。また、例として挙げた気温と売り上げだけのデータを見ても、気温が高くなると売り上げの上がり方が急になっている部分があります。そこで、実際には気温 x と売り上げ y の関係を、直線ではなく、より複雑な曲線で予測の方が適切かもしれません。その場合は最小化すべき関数も2次関数ではなく複雑なものになります。

ここまでの話を総合すると、多くの変数をもつ複雑な関数の最小について調べることでさらに実践的なデータサイエンスが実現でき、そのために必要な高度な数学は数理最適化という分野で研究されています。ただ、この回帰分析の話はあくまで一例で、実際のデータサイエンスでは数理最適化以外にもさまざまな数学が活躍します。そこで、数学のいろいろな分野を基礎からしっかりと学ぶことがとても重要です。

【図1】 気温とアイスクリームの売上の関係の例



【図2】 最小化したい関数のグラフ



【図3】 最小化したい関数の等高線

