

# もうひとつの汎帰納推論機械：PLRMとdPLRM

田邊 國士 統計思考院 特任教授

画像診断などの医学分野で近年急速に応用が進んだ深層学習は非常に多くの訓練用データサンプルを必要とするデータハングリーな方法である。しかし一般にサンプル数が少ない医学分野のデータに深層学習が適しているとはいえず、サポートベクターマシンやランダムフォレストなどの方法が有効であることも多い。私達（山梨大と島津製作所との共同研究）は少数サンプルの場合にも有効なベイズモデルに基づく学習機械である罰金付ロジスティック・リグレーション・マシンPLRMとその双対機械 dPLRM (Tanabe2001). を、質量分析データに基づくがん診断装置の開発に応用した

物理現象とは異なり生体にかかる現象は複雑多様な動態を呈する。対象が単純な「もの」である物理現象とは違い、異なる時間スケールで変化する多様な因子が階層的に結合した多自由度を持つシステムとして取り扱う必要があり、医療や生物学においては機械学習が従来の仮説演繹法に勝る有効な方法論を与える。このとき生体現象を観測・測定するに当たって観測対象をどう分節化するかは自明では無い。言い換えれば何を因子と見なすかは主観的である。機械学習の最も重要な特長は、主観的な分節化を排除した素データからも有効な推論を可能とする事である。また、マルチモーダルな生データの解析への適合性が高い。

機械学習の本質は因果関係の先験的想定する従来の推論法と異なり、その想定無しに事象の分節化に先立って観測できる現象のデータ変数間の共起確率を捉える方法であると言える。従って確率統計モデルに基づく推論が自然な接近法である。深層学習も近年ベイズモデル化の方向に発展している。しかし深層学習には解釈不能性の問題がある。罰金付ロジスティック・リグレーション・マシンにおいては主・双対機械を同時に運用できるのでこの点で少々有利である。

### 1. がん診断

がん診断において特定のがんに特異的に検出されるバイオマーカーと呼ばれる生化学因子の測定がおこなわれている。昨今ニュートン-デカルト・パラダイムに沿って各種がんに対応するバイオマーカーの発見に多くの資源・人材が投入されている。しかし、個々のがんにたいして1, 2のバイオマーカーを探索することは賢明なことであろうか？ がんは遺伝子上の突然変異によって引き起こされる。たとえば急性骨髄性白血病と診断される病態も遺伝子上の変異の仕方は多種多様である。細胞内外において相互作用をする数多くの生化学分子群の代謝経路における分子群の分布もこの変異を受けて非常に複雑・多様な動態を呈する。一般にがんの病名はがん細胞の形態学的観察および臨床的所見に基づく命名に過ぎず、実体的に定義されるものではない。したがってそれらの代謝物質のひとつふたつをバイオマーカーとして特定する

ことに努めるよりも、出来る限り広い範囲の代謝分子群の発現を観測して、範囲内のすべての分子の発現およびがん・非がんの共起関係に基づいて診断をする方が理に適っている。

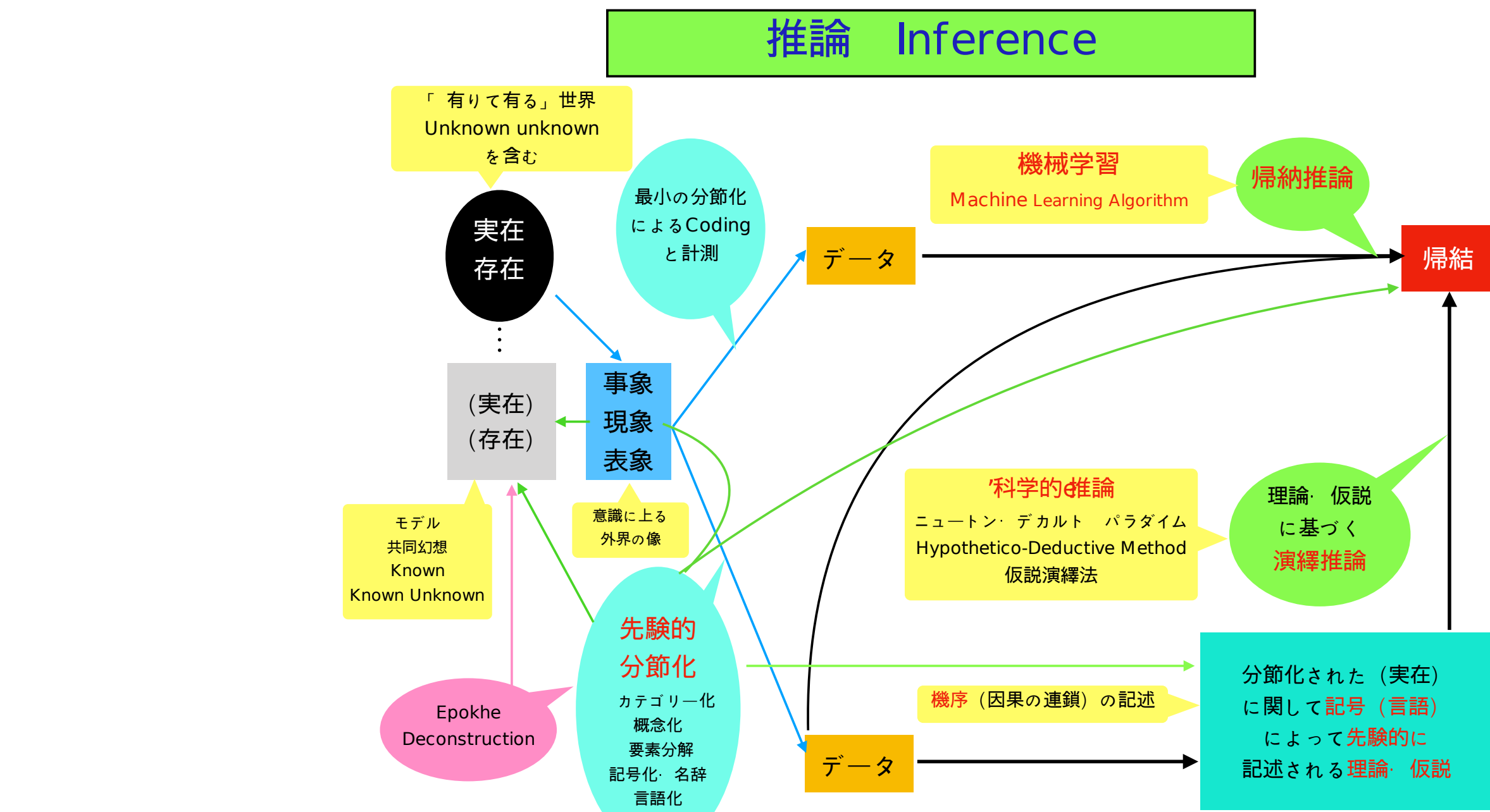
私達はこの点に着目して、細胞片から採取した極微量の液滴に含まれる代謝化合物の質量分析データに帰納推論(学習)機械 dPLRMを適用してがん診断支援装置を開発した。質量分析装置は一次的にはイオン化分子のスペクトルが輻湊した生データを出力する。この生データの情報を基にデータベースとソフトウェアを用いて主要な含有分子を同定するのが一般的な分析装置の利用法である。ニュートン-デカルト・パラダイムの方法論は、対象因子を分子として分節化することを要求するからである。しかし興味深いことに、われわれの診断支援装置においては質量分析器が一次的に出力する生の測定データを入力して診断に成功している。しかも質量分析装置の分解能の1000倍にくらいの中でのbinningしたデータを用いて訓練したにも関わらず良い成績を収めている。機械学習に基づく診断には分子同定という特徴量の先験的な分節化や精密な測定はを必ずしも要しないのである。

### 2. PLRMとその双対機械dPLRM

ベアとなるこれらの学習機械の詳細は参考文献に譲るが、PLRMは過完備なフレームと呼ばれる一次独立性を欠く大量の非線形回帰関数族を用いるロジスティック回帰モデルに回帰係数に事前分布を導入したモデルで、学習に必要な最適化数値計算をニュートン法アルゴリズムによって実行できる点に長所がある。さらにこの最適化モデルの双対問題はいわゆるカーネル回帰に数学的に帰着し、やはりニュートン法アルゴリズムが適用できる (Tanabe2001). 後年、Gaussian Processモデルなどが多用されるようになったが、主・双対の2つの関係性が閑却されがちである。昨今双対モデルに対応するKernel法が脚光を浴びているが、Kernel法の裏側にある主モデルを同時に考察するべきであると私は考えている。このことは解釈可能性に問題がある機械学習に対する解決の鍵にもなり得る事実である。

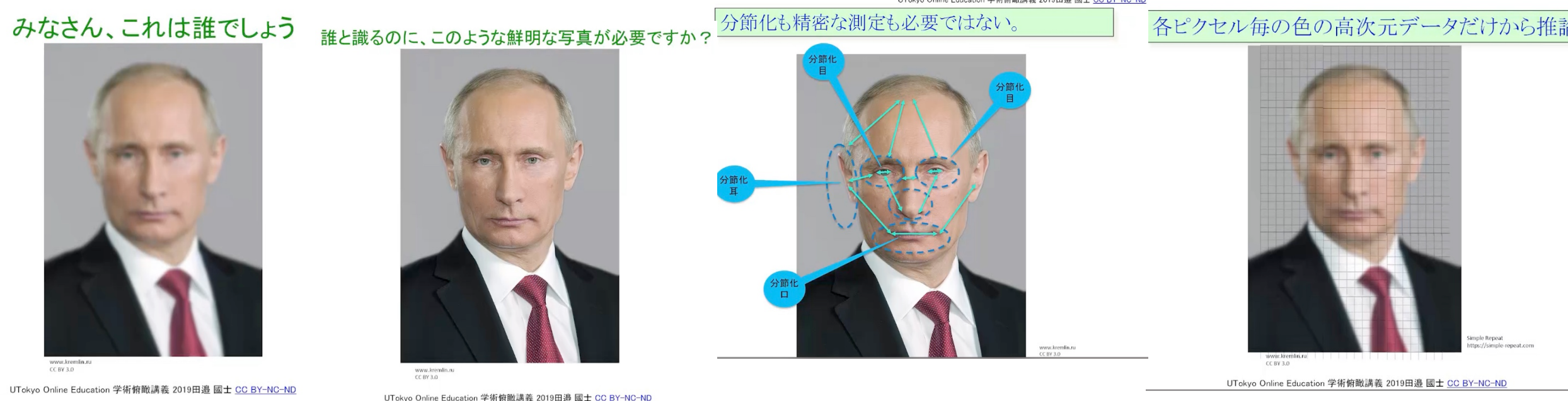
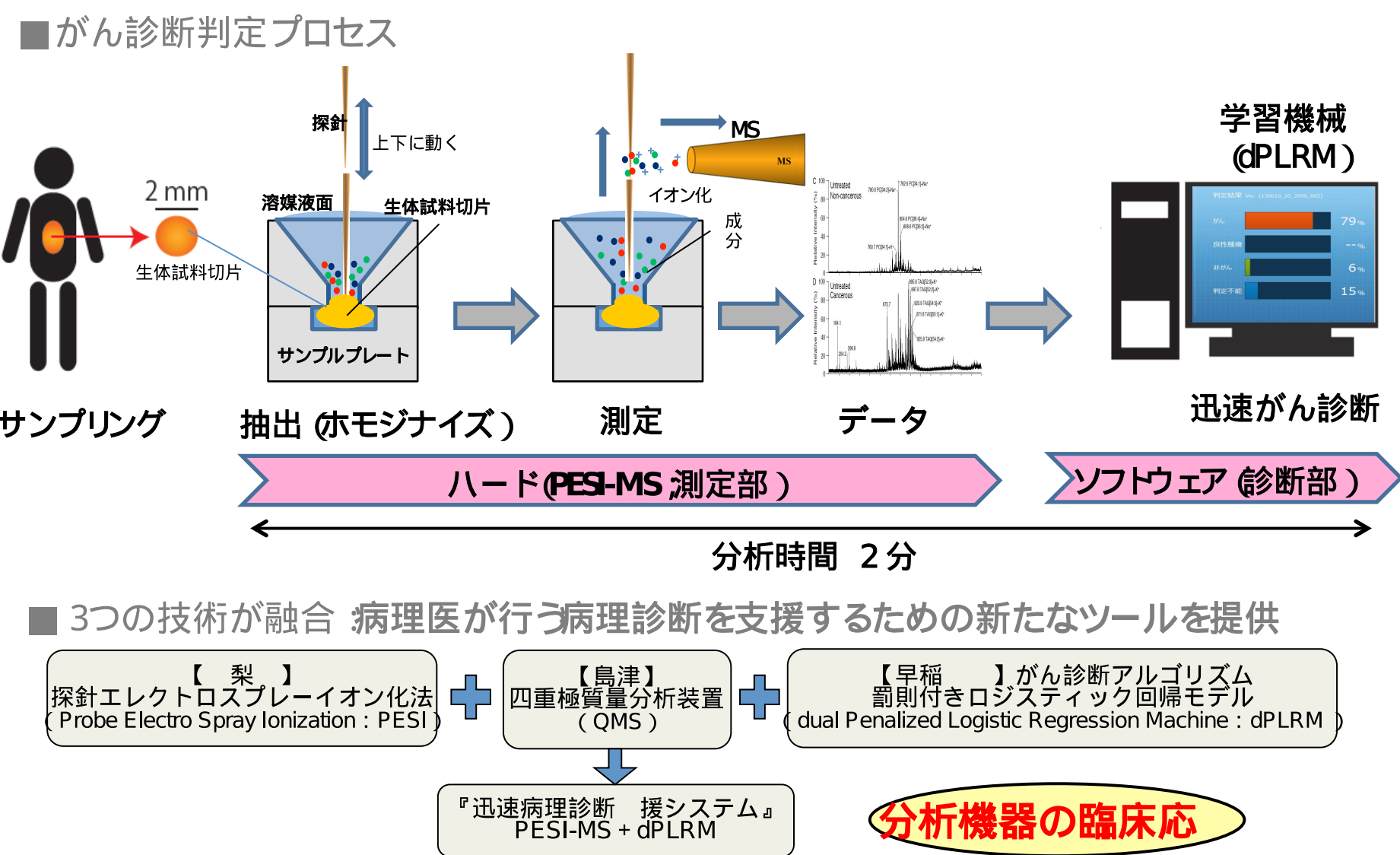
### 3. 回帰モデルを超えて共起（同時）分布モデルパラダイムへ

帰納推論機械の本質は共起確率のコンパクトな表現モデルにある。DNNにしてもdPLRMにしても多くの機械学習モデルは回帰モデルである。確かに回帰モデルも条件付き確率を通して共起確率を表現するひとつの方法ではある。しかし帰納推論にはもっと別の道があると思われる。因みに私は学習機械を汎帰納推論機械と呼んでいるが、それは学習するのではなく共起確率分布を創発する機械であると考えからである。。



竹田扇先生 (山梨大)作成スライド

### 迅速病理診断支援システム 概要



## References

- [1] Tanabe K. , Penalized Logistic Regression Machines: New methods for statistical prediction 1. Institute of Statistical Mathematics, Cooperative Research Report 143, 2001 163-194.
- [2] Tanabe K., Penalized Logistic Regression Machines: New methods for statistical prediction 2. Proceedings of 2001 Workshop on Information-Based Induction Science (IBIS2001) 2001:71-76.
- [3] Tanabe K. , Penalized Logistic Regression Machines and Related Linear Numerical Algebra. 京都大学数理解析研究所講究録 2003; 1320; 239-249.
- [4] 田邊國士, 帰納推論機械 PLRM と dPLRM ー方法論, モデル, アルゴリズムおよび応用. システム・制御・情報 2007; 51(2):87-95
- [5] 田邊國士, ポスト近代科学としての統計科学. 伊庭幸人編, ベイズモデリングの世界, 岩波書店 2018: 87-105.
- [6] 田邊國士, 帰納推論と経験ベイズ法, 甘利俊一他編, 階層ベイズモデルとその周辺, 岩波書店 2004: 236-252.
- [7] Yoshimura K. et al. , Analysis asof Renal Cell Carcinoma as a First Step for Developing Mass Spectrometry-based Diagnostics, Journal of the American Society for Mass Spectrometry, 2012, 23(10):1741-9.
- [8] Tanabe K et al., dPLRM Revisited: Towards Cancer Diagnosis with PESI-Mass Spectrometry Data. Poster presented at the 61st Annual Conference on Mass Spectrometry, Mass Spectrometry Society of Japan 2013.
- [9] Yoshimura K. et al., Real-time diagnosis of chemically induced hepatocellular carcinoma using a novel mass spectrometry-based technique, Analytical Biochemistry, 2013, 441:32-37.
- [10] Yamagishi T et al., 608 Probe Electrospray Ionization-Mass Spectrometry and Bayesian Statistics: A Potential of novel cancer Diagnostics System in Renal Cell Carcinoma. The Journal of Urology 2013; 189(4S)Supplement: e248-249.
- [11] 竹田扇他, 質量分析法と統計的学習機械を組み合わせた新規がん診断支援装置の開発. 島津評論 2013; 69(3・4):203-210.
- [12] 竹田扇他, 分析と総合: 相反する視座からの新しいがん診断支援装置の試み, 臨床病理 2015; 63(9):1083-1089.
- [13] 吉村健太郎, 癌の高精度迅速診断システムの開発を目指して. 癌と人 2014; 41: 58-58.
- [14] 出水秀明他, 質量分析装置, 及び該装置を用いた癌診断装置. 特許6189587号, 2017.
- [15] 田邊國士, 統計的学習機械の医療応用 — PLRM,dPLRMお場合, 医学のあゆみ, 2020, 274(9): 718-723.



大学共同利用機関法人 情報・システム研究機構

統計数理研究所

The Institute of Statistical Mathematics