

09

機械学習と特徴量

ここでは、機械学習がどのような概念なのかをもう少し深く掘り下げていきます。機械が知能を持つとはそもそも何を意味するのか、同時に、その上で極めて重要な概念である特徴量についても、解説します。

○ 機械が知能を持つこと

機械が知能を持つとは、ものごとを「分ける」ことができるようになることといえます。たとえば今日はアイスが売れるのか、売れないのか。ある物体がリンゴなのか、そうではないのか。ある事業で利益が出るのか、出ないのか。

Section04で学んだように、人工知能はパターン探索や知識の蓄積でこの「わかる」機能を実現しようとしてきましたが、どちらもパターンが多くなるなどの問題でうまくいかなくなっていました。

機械学習とは、この「分ける」能力の実現のために、今までの「演繹的思考」ではなく「統計的思考」という新たなアプローチを取り入れることで、知能を実現しようとする人工知能です。ここでの演繹的思考とは「AはBである」といった根拠から思考をおこなうことであり、対して統計的思考とは「AはBである確率が高い」といった根拠から思考をおこなうことを指します。

■ 演繹的思考と統計的思考

演繹的思考

- ① 気温の高い日はアイスが売れる
- ② 今日は気温が高い

③ 今日はアイスが売れる

統計的思考

データ	気温	売り上げ
7/1	25℃	△
8/1	32℃	○
9/1	30℃	○
10/1	18℃	×
⋮	⋮	⋮

気温28℃以上の日はアイスが売れる

統計は、統計処理に耐えうるデータが存在しなければ成り立ちません。昔から統計的なアプローチに関する研究はありましたが、こんにち機械学習が一世を風靡するようになったのは、コンピュータやインターネットの普及により、大量のデータとそれを処理する計算資源が手軽に用意できるようになったからです。

この統計的思考というアプローチによって高度な知能を獲得した機械学習ですが、まだ弱点があります。それは、機械学習がデータを取り込む方法によるものです。

ものを見分けるとき、人間は見た目や匂い、感触などから情報を入手して見分けます。対して機械学習は、情報を果物の色の濃淡や重さ、匂い成分の量などの「**特徴量**」という数値で取り込みます。この特徴量を決めるのは人間の仕事であり、このことを**特徴量設計**といいますが、実はこの特徴量の決め方でアルゴリズムの性能は大きく変わってしまうのです。たとえばリンゴとナシを見分ける特徴量として、「赤さ」や「甘さ」などはよさそうですが、「丸さ」や「表面のなめらかさ」はあまり差がなさそうだと分かります。

■ よい特徴量とは

	赤さ	甘さ	丸さ	なめらかさ
リンゴ1	0.9	0.6	0.91	0.1
リンゴ2	0.95	0.55	0.92	0.2
リンゴ3	0.92	0.59	0.89	0.1
ナシ1	0.21	0.8	0.88	0.2
ナシ2	0.17	0.9	0.9	0.1
ナシ3	0.2	0.95	0.95	0.2

特徴量に大きな差がある

よい特徴量

特徴量に大きな差がない

よい特徴量ではない