最適化理論

オリエンテーション

統計的機械学習

データから対象の未知の構造を見いだしたり、学習した結果をもとにして予測を行う機械 (=アルゴリズム）の構成を目指す。

- 機械学習問題への数理的アプローチ
- 現在、非常に活発に研究が進められている分野
- 重要なキーワード：<u>最適化と確率</u>

統計的機械学習の技法

- 線形回帰モデル、深層ニューラルネットワーク、関数フィッティング、ベイズ推論、サポートベクトルマシン、ブースティング、グラフィカルモデル、ビリーフプロパゲーション、変分推論、MCMC

統計的機械学習の応用先

- パターン認識、信号処理、データマイニング、遺伝子データ解析、デジタル通信工学、人工知能、気象予測、ロボティクス、画像処理。。。

- 計算機の高速化により巨大なデータセットが扱えるようになりつつある。
- 複雑なアルゴリズムも実装可能。

講義の概要

本講義では、「確率と最適化」に焦点を当てて最近の機械学習技術の理解に必要となる基礎を身につけることを目指す。

- 確率計算の基礎とベイズ推論
- 事後確率計算
- 最尤推定・最大事後確率推定
- ベイジアンネットワーク
- 関数近似と回帰
- 深層ニューラルネットワーク入門
- 凸計画問題
- 勾配法
- 双対問題

「知的」なアルゴリズムが持つ３要素

- **表現**
- **推論**
- **学習**

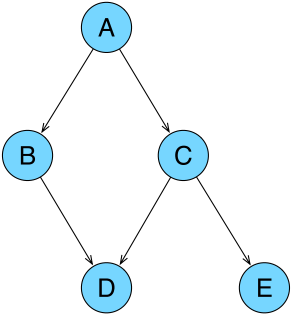
表現: 確率モデル

対象とする系 (システム) を相互に確率的に相関を持つ確率変数の集まりとしてモデル化する (確率モデル)。

- **今日の円・ドル為替レートを確率変数 X で表す。**
- **明日の T 社の株価を確率変数 Y で表す。**
- **同時分布 $P_{XY}(x, y)$: 確率モデル I**
- **条件付分布 $P_{Y|X}(y|x)$: 確率モデル II**

表現: グラフィカルモデル

ベイジアンネットワーク



- 確率変数をノードとする。
- 確率変数間の確率的依存関係をエッジとする。

推論: 事後確率分布の計算

- 我々の文脈では、「推論」とは、興味のある確率変数に対して、事後確率分布を計算すること である（ベイズ的考え方）。
- 推論者 (推論器) は、系に対する知識 として確率モデル (同時分布または条件付分布) を持つ。
- 推論者は、得られた観測値から 事後確率分布を計算する。

推論: 典型的な推論過程 (2 変数の場合)

- X, Y: 確率変数
- 推論者は、事前確率分布 $P_X(x)$ と条件付確率分布 $P_{Y|X}(y|x)$ を知っている (確率モデル)
- いま、推論者は、確率変数 Y の実現値 y^* を観測した！
- 推論者の仕事は、事後確率分布 $P_{X|Y}(x|y^*)$ を計算すること。
- 知識の更新を行う: $P_{X|Y}(x|y^*) \rightarrow P_X(x)$

ベイズ則を使う

観測値
(observation)

尤度関数
(likelihood function)

事前確率
(prior probability)

$$P_{X|Y}(x|y^*) = \frac{P_{Y|X}(y^*|x)P_X(x)}{P_Y(y^*)} = \frac{P_{Y|X}(y^*|x)P_X(x)}{\sum_x P_{Y|X}(y^*|x)P_X(x)}$$

事後確率分布
(posterior probability)

正規化定数
(normalization constant)

ベイズ推論の枠組み

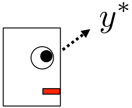
(1)

推論器

初期状態として
事前確率分布 $P_X(x)$
と
条件付確率分布 $P_{Y|X}(y|x)$
を持つ

(2)

観測値

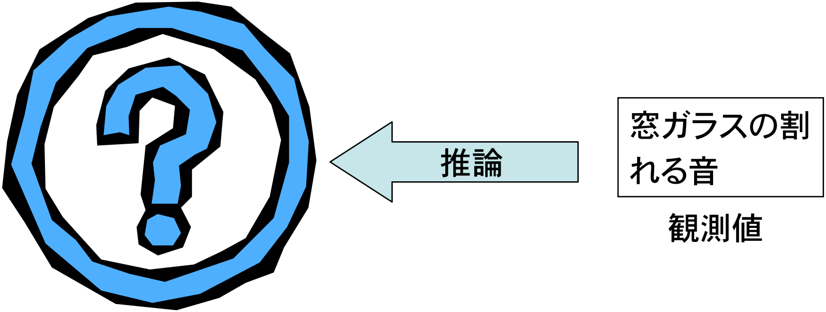


観測値から
事後確率分布 $P_{X|Y}(x|y^*)$
を計算する。
この事後確率分布
に基づき様々な処理を
行う。

- **推論結果 $\longrightarrow$ 事後確率分布**
- **推論 $\longrightarrow$ 事後確率分布の計算**
- **確率の形で知識を蓄え、推論を行い、知識を更新していく**

具体例

- 夜中に寝ていると自宅の窓ガラスが割れる音に気がついた。



原因の可能性

- 泥棒が侵入
- 隕石が衝突
- 敵が攻めてきた
- 撃たれた
- その他。。。

If-then 形式の知識に基づく推論

知識

- **もし、隕石が窓に衝突すると窓が割れる**
- **もし、窓が銃撃されると窓が割れる**
- **もし。。。**

推論においては、Then **節から逆に辿る：窓が割れた→隕石** or **銃撃** or …

- **このアプローチの欠点** $\rightarrow$ **原因のもっともらしさが分からない**

確率に基づく推論

- 隕石が落ちてくる確率 0.00001
- 隕石が当たると窓が割れる確率 0.9
- 銃撃される確率 0.01
- 銃撃されると窓が割れる確率 0.99

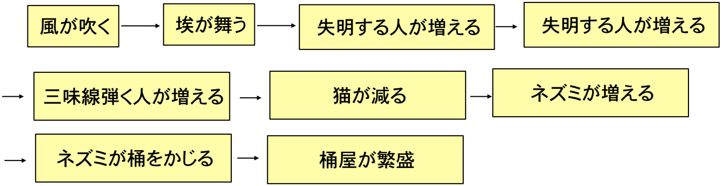
観測値である「窓が割れた音」から原因の事後確率分布を計算

- 原因の「もっともらしさ」を定量的に評価できる。

ことわざで

「風が吹けば桶屋が儲かる」
というのがありますが知っていますか？

確率的な因果関係



現実的な因果関係は不確実性を含む

- 観測情報は、なんらかの「雑音」の影響を被る
- 因果関係に関する我々の知識は不完全
- そもそも「因果関係自身が確率的」な場合もある

確率モデルで定式化

- X: **風が吹いたか否か** $(0/1)$
- Y: **桶屋が儲かったかどうか** $(0/1)$
- $P_X(x)$: **風が吹く事象に関する確率分布**
- $P_{Y|X}(y|x)$: **風が吹いて桶屋が儲かるかどうかの確率モデル**

$$
\begin{aligned}
P_{Y|X}(0|0) &= 0.9 \ (\text{風が無く儲からない確率}) \\
P_{Y|X}(1|0) &= 0.1 \ (\text{風が無く儲かる確率}) \\
P_{Y|X}(0|1) &= 0.05 \ (\text{風が吹いても儲からない確率}) \\
P_{Y|X}(1|1) &= 0.95 \ (\text{風が吹いて儲かる確率})
\end{aligned}
$$

$y^* = 1$(**桶屋が儲かった）のを見て、風に関する事後確率分布** $P_{X|Y}(x|y^*)$ **を推論するのが確率推論。** $y^* = 1$ **から、風が吹いたか、否かを** $0, 1$ **で推論するのではないことに注意したい。**

2 変数の場合：ベイズ則ふたたび

観測値
(observation)

尤度関数
(likelihood function)

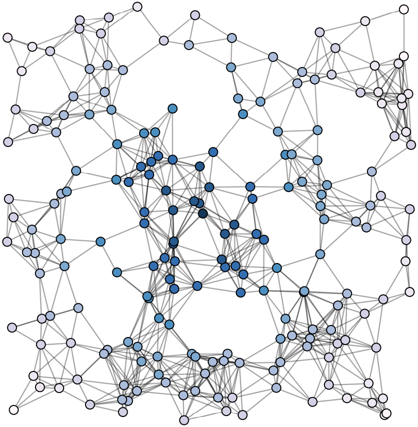
事前確率
(prior probability)

$$P_{X|Y}(x|y^*) = \frac{P_{Y|X}(y^*|x)P_X(x)}{P_Y(y^*)} = \frac{P_{Y|X}(y^*|x)P_X(x)}{\sum_x P_{Y|X}(y^*|x)P_X(x)}$$

事後確率分布
(posterior probability)

正規化定数
(normalization constant)

こんな場合は？



なぜ「確率」に基づく推論を学ぶ必要があるのか

- 「ベイズ的視点・考え方」は、科学・工学・ビジネス分野を問わず、不確実性を取り扱う必要のある現代人にとって<u>常識とすべき考え方</u> である。
- グラフィカルモデルに関する知識は、より広く・深く機械学習を学ぶ上必ず身に付けておくべき
- 実世界にたくさんの応用事例がある。
- 深層ニューラルネットワークの世界を深く学んでいくために必要
- 物事を考える上で新しい視点を持つことができる。

では、なぜ「最適化」を学ぶ必要があるのか

多くの統計的機械学習の手法では、連続最適化が「モデル学習のための基礎技術」となっている。特に、連続最適化（特に凸最適化）の基礎は、統計的機械学習を学ぶときには必須の知識となる。

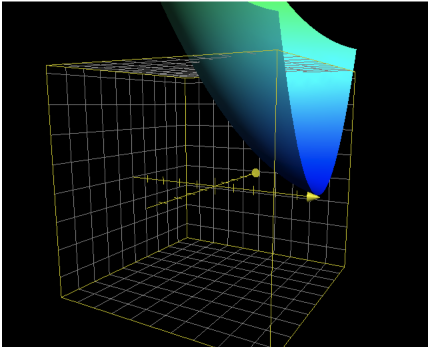
- 推論のための柱：確率
- 学習のための柱：最適化

2 次計画問題の例

$$\text{minimize } (x_1 - 6)^2 + 2x_2^2$$

subject to

$$\begin{aligned} x_1 &\leq 5 \\ -x_1 &\leq 5 \\ x_2 &\leq 5 \\ -x_2 &\leq 5 \end{aligned}$$



SVM の学習 $=2$ 次計画問題

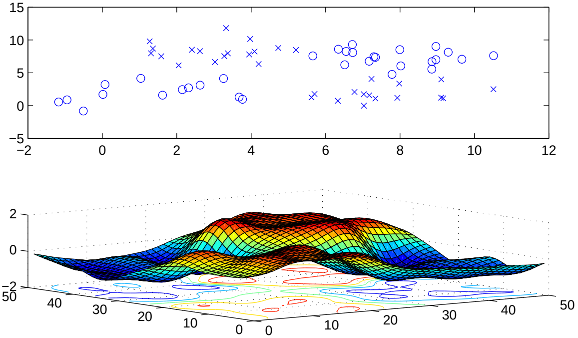
$$\text{minimize } \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\lambda_i\lambda_j t_i t_j K(x_i, x_j) - \sum_{i=1}^{m}\lambda_i$$

subject to

$$\lambda \geq 0, \quad \sum_{i=1}^{m}\lambda_i t_i = 0$$

ここで、$K(x_i, x_j) = \phi(x_i)\phi(x_j)^T$ **である。**

サポートベクトルマシンの出力



講義の概要

本講義では、「確率と最適化」に焦点を当てて最近の機械学習技術の理解に必要となる基礎を身につけることを目指す。

- 確率計算の基礎とベイズ推論
- 事後確率計算
- 最尤推定・最大事後確率推定
- ベイジアンネットワーク
- 関数近似と回帰
- 深層ニューラルネットワーク入門
- 凸計画問題
- 勾配法
- 双対問題

講義の進行

- 最初に講義のスライドの説明を聞きます (40–60 分)
- 演習問題を解きます。
- 解答の解説を理解します。

途中でパソコンを使った演習 (深層ニューラルネットワークのあたり）も入れる予定です。