言語処理のための機械学習入門

1 .必要な数学的知識

1.1 準備と本書における約束事

1.2 最適化問題

河野　和平

1.1 準備と本書における約束事

単語分割

- 与えられた文を単語に分割するタスク

「私の名前は山田花子です。」

→　私　の　名前　は　山田　花子　です　。

（英文の場合）

単語間に空白が設けられるため不要。

「My name is Hanako Yamada.」

品詞タグ付け

- 文中の単語の品詞を推定するタスク

「ここにリンゴがある。」

→　(ある)は動詞

「ある人と一緒に出かけた。」

→　(ある)は連体詞

同じ(ある)でも品詞が異なる。

高精度な品詞タグ付けは容易でない。

構文解析・文書分類

構文解析

- 文の構文的な構造を推定するタスク

私は　海へ　行く。

文書分類

- 与えられた文書をいくつかのクラスのうち適切なクラスに決定する。

機械学習

- コーパス

言語の様々な形の用例の集まり。

- 機械学習を用いた言語処理

コーパスから様々な数値を取りだし、 数学的な知識を用いて何らかの処理を加える。

1.2 最適化問題

最適化問題

- ある制約のもとで関数を最大化(最少化)する変数値とその時の関数値を求める問題。
- 最大化問題、最少化問題とも呼ばれる。

$$
\begin{array}{lll}
max. & f(x) & \cdots\text{目的関数} \\
s.t. & g(x) \geq 0 & \cdots\text{不等式制約} \\
 & h(x) = 0 & \cdots\text{等式制約}
\end{array}
$$

最適化問題

- 制約式を満たす解のことを<u>実行可能解</u>と呼ぶ。
- 実行可能解の集合を<u>実行可能領域</u>という。
- $max.\ \ f(x)$ と $min.\ \ -f(x)$ は等価であり、本質的な違いはない。

凸計画問題

- 目的関数の値が改善する方向に進んでいけば、解にたどりつく問題。
- 比較的解きやすい問題であり、解法はある程度確立している。

解きたい問題を凸計画問題で表わすことができれば、問題解決に大きく近付く

凸集合

凸集合　　　　凸でない集合

- 任意の $x^{(1)}, x^{(2)} \in A$ があったとき任意の$t \in [0,1]$ に対して

$$t\,x^{(1)} + \left(1 - t\right) x^{(2)} \in A$$

が成り立つとき、集合A は凸集合であるという。

集合内の任意の点を結ぶ線分が
集合自身からはみ出ないことを表す。

凸関数

- 凹関数(上に凸な関数) $f(x) = -x^2$

任意の $x^{(1)}, x^{(2)} \in R^d$, $t \in [0,1]$ に対し、

$$f\left(t\,x^{(1)} + (1-t)\,x^{(2)}\right) \geq t f\left(x^{(1)}\right) + (1-t)\,f\left(x^{(2)}\right)$$

が成立する。

- 凸関数(下に凸な関数) $f(x) = x^2$

上下が逆なだけで性質は同じ

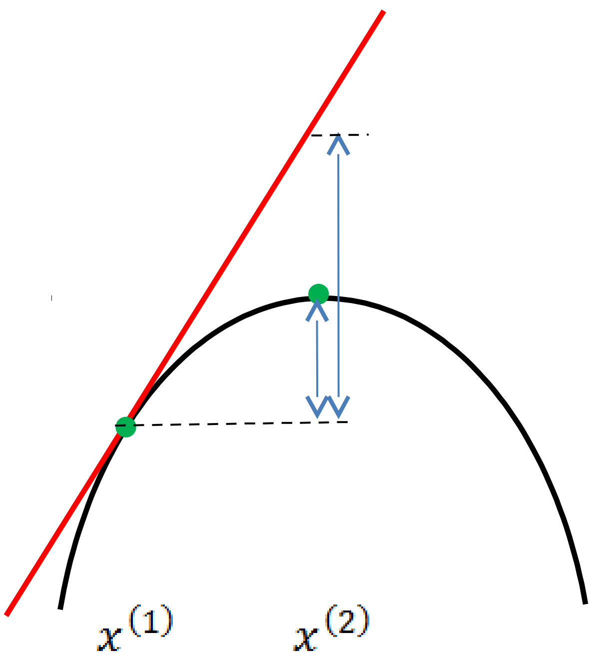
凸関数

- 上に凸な2つの関数の和は上に凸な関数である。
- $f(x)$ が上に凸な関数ならば a が非負の定数のとき、 $af(x)$ も上に凸な関数である。

凸関数

- 凹関数(上に凸な関数)であるための１次の条件
  - すべての接線がその関数の上側に来る

$$f\left(x^{(2)}\right) - f\left(x^{(1)}\right) \leq \frac{\partial f\left(x^{(1)}\right)}{\partial x}\left(x^{(2)} - x^{(1)}\right)$$



凸関数

- 凹関数(上に凸な関数)であるための２次の条件
  - ２階微分 $f''(x)$ は常に0以下となる。

1次の条件より

$$f(x+\delta x) - f(x) \leq \frac{\partial f(x)}{\partial x}\delta x$$

テイラー展開より

$$\begin{aligned} & f(x+\delta x) - f(x) \\ &= f(x) + f'(x)\delta x + \frac{f''(x)}{2}\delta x^2 + \mathrm{O}(\delta x^3) - f(x) \\ &= f'(x)\delta x + \frac{f''(x)}{2}\delta x^2 + \mathrm{O}(\delta x^3) \end{aligned}$$

凸関数

- 凹関数(上に凸な関数)であるための２次の条件

$$\frac{f''(x)}{2}\delta x^2 + O(\delta x^3) \le 0$$

従って、 $$f''(x) \le 0$$

- 多変数関数の場合

(i , j)成分を $H_x = \frac{\partial f(x)}{\partial x_i \partial x_j}$ で表わすヘッセ行列が

$x^T H_x x \le 0$ (半負定値) を満たす。

凸計画問題

- ある最適化問題において
  - 目的関数が凸関数であり、かつ実行可能領域が凸集合であるとき、<u>凸計画問題</u>と呼ぶ。

- 変数の取り得る値について制約がない場合
  - 目的関数は凸関数であるので、微分して0になる点を求めればよい。

凸計画問題

- 数値解法

目的関数:　$f(x)=\frac{1}{2}x^2+exp(x)$

微分:　$f'(x)=x+exp(x)$

$$x+exp(x)=0$$

閉形式(x ＝の形)で解が求まらない。
なんらかの解を初期値として与え、
より良い解へと値を更新し、最適解を求める。

最急勾配法

- 目的関数の傾きが最も急な方向に変数の値を変化させていく方法。

最急上昇法

最大化問題なら目的関数が最も増加する方向

$$x^{new} = x^{old} + \epsilon \nabla_x f(x)$$

最急降下法

最少化問題なら目的関数が最も減少する方向

ニュートン法

- 最大化問題

$$x^{new} = x^{old} + \epsilon H_{x^{old}}^{-1} \nabla_x f(x)$$

  - ヘッセ行列を用いて x の更新を行う。
  - ニュートン法では目的関数の2階微分まで考慮する。
  - 逆行列の計算は大変なため、高速近似法を用いることが多い。

等式制約付凸計画問題

$$\begin{aligned} &max. \quad f(x) \\ &s.t. \quad g(x) = 0 \end{aligned}$$

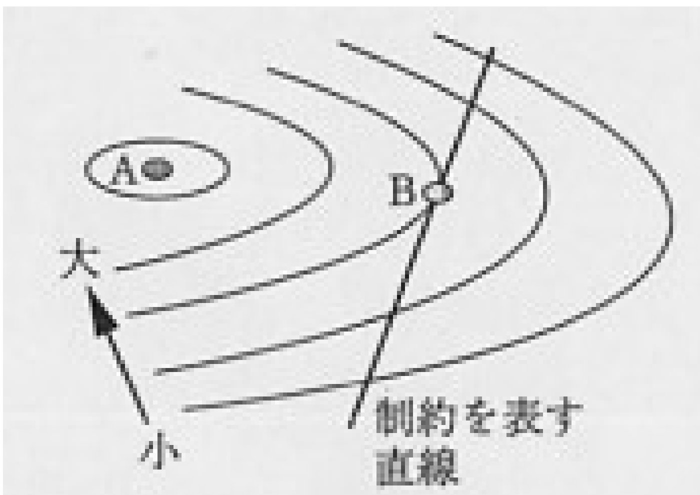
- 目的関数の偏微分が0になる点が制約条件を満たしているとは限らない。

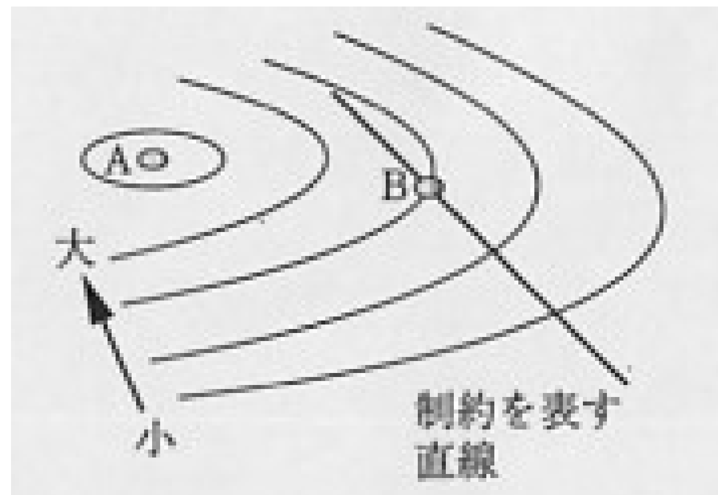
- ラグランジュ関数　$L(x, \lambda) = f(x) + \lambda g(x)$

  ラグランジュ関数の x に関する偏微分が0となり、かつ与えられた制約が満たされるとき、最適解が得られる。

ラグランジュの乗数法

- xが2次元空間に存在する場合





ラグランジュの乗数法

- 制約が複数の場合

  それぞれの制約 $g_i(x) = 0$ に対して、一つのラグランジュ乗数 λ_i を与える。

$$L(x,\lambda) = f(x) + \sum_i \lambda_i g_i(x)$$

不等式制約付凸計画問題

$$\begin{aligned} max. \quad & f(x) \\ s.t. \quad & g(x) \geq 0 \end{aligned}$$

- ラグランジュ関数 $L(x,\lambda) = f(x) + \lambda g(x)$

  ただし、$\lambda \geq 0$ であるとする。

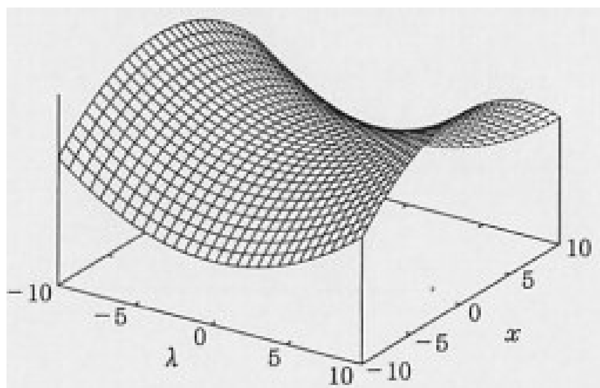
  制約を満たす x において、$\lambda g(x) \geq 0$ であり、
  $L(x,\lambda)$ はより大きくなる。

鞍点

- ラグランジュ関数 $L(x,\lambda)$ について、

$$L\left(x,\lambda^*\right) \le L\left(x^*,\lambda^*\right) \le L(x^*,\lambda)$$

を満たすような点 $\left(x^*,\lambda^*\right)$ を鞍点と呼ぶ。



x に関して最大値となり、

λ に関して最小値となるような点。

鞍点

- 凸計画問題において、鞍点は最適解を与える。

①主問題

$L(x,\lambda)$ を最大化する x を求める。

そのような x は λ に関する式 $x^*(\lambda)$ で表される。

②双対問題

$L(x^*(\lambda),\lambda)$ を λ について最少化する。

λ^* が求まれば、最適解 x^* も得られる。