

## 内容

1.	線形回帰モデル.....	2
1. 1	線形回帰.....	2
1. 2	正則化.....	2
2.	非線形回帰モデル.....	3
3.	ロジスティック回帰モデル.....	4
4.	主成分分析.....	5
4. 1	解説.....	5
4. 2	問題.....	5
5.	アルゴリズム.....	6
5. 1	K 近傍法 (KNN) .....	6
5. 1	K 平均法 (K-means) .....	7
6.	サポートベクターマシン.....	8
6. 1	解説.....	8
6. 2	問題.....	8

## 1. 線形回帰モデル

### 1. 1 線形回帰

1) 線形回帰は統計でも用いられる最もシンプルなモデルの1つ。

データ（分布）が有った時に、そのデータに最もあてはまる直線を考えるということ。

2) 線形回帰は、説明変数と目的変数の関係に、直線や（超）平面を当てはめ、予測・説明する教師あり学習の代表的な手法である。

$$y = \alpha + \beta x$$

・説明変数の数が複雑になると

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots$$

・直線から平面になる。

・3次元以上の平面を超平面という。

・（偏）回帰係数の仮説検定

「（偏）回帰係数に意味がある」とは、その説明変数が目的変数を説明するのに実際に寄与していること

3) 単回帰分析・重回帰分析

### 1. 2 正則化

1) ラッソ回帰

2) リッジ回帰

## 2. 非線形回帰モデル

非線形回帰では、連続応答変数と 1 つ以上の予測変数との間の非線形な関係を表す方程式を生成し、新しい観測値を予測する。

$$y = f(\mathbf{X}, \beta) + \varepsilon$$

$\beta$  : 計算される非線形パラメータの推定値

$\varepsilon$  : 誤差項

### 3. ロジスティック回帰モデル

ロジスティック回帰は2クラス分類に用いられる。

#### 1) 2クラス分類

出力にシグモイド関数を使用する。

任意の値を0から1の間に写像するシグモイド関数を使用することによって、与えられたデータを正例(+1)になるか、負例(0)になるかの確率が求められる。

#### 2) 多クラス分類

出力にソフトマックス関数を使用する。

#### 3) 問題1

シグモイド関数により、出力を(0, 1)の範囲に制限することで、モデルの出力を確率値  $p(y=1|x)$  で扱えるようにしている。

$n$  個目のデータに対する目的変数とモデルの出力をそれぞれ  $y$ 、 $\hat{y}$  としたとき、ロジスティック回帰では全データにおける負の対数尤度関数

$$-\sum (y_n \log \hat{y}_n) + (1 - y_n) \log (1 - \hat{y}_n)$$

を最小にするようなパラメータ  $w, b$  を学習する。

学習により得られたパラメータの解釈にはオッズが用いられる

$$\text{オッズ} = \hat{y} / (1 - \hat{y}) = \exp(w^T x + b)$$

$$p(y=1|x) / p(y=0|x)$$

$$\hat{y} = 1 / (1 + \exp(-w^T x - b))$$

#### 4) 問題2

ロジスティック回帰は分類問題を解くための手法であり、一般化線形モデルの一種である。

例えばマーケティングにおける適用としては、見込顧客が購買行動に至る確率を予測することが挙げられる。

ある事象が起こる確率  $p$  と起こらない確率  $(1 - p)$  の比の対数、つまり対数オッズを線形回帰するためである

$$\text{数式} = \log(p / (1 - p))$$

## 4. 主成分分析

### 4. 1 解説

- 1) データのバラつきを最も説明できる順番で直行する軸を抽出する手法
- 2) データの特徴量間の関係性、すなわち相関を分析することでデータの構造をつかむ手法。  
特に特徴量の数が多い場合に用いられ、相関を持つ多数の特徴量から、相関の少ない少数の特徴量へと次元削減すること。
- 3) 寄与率の大きい順番に、第一主成分、第二主成分・・・と呼ぶ
- 4) カイザー基準では、固有値が1以上である主成分を求める。
- 5) スクリー基準では、スクリープロットの推移が滑らかになる直線までを選択
- 6) 平行分析：もとのデータと同じ大きさの正規乱数行列の固有値と比較する。
- 7) 複数の変量を相関のない少数で全体のバラつきを最もよく表す主成分と呼ばれる変量を合成する手法
- 8) 類似する複数の説明変数を主成分として、**合成**することで次元削減を行う。
- 9) 因子分析：変量間に存在する潜在要因を探し出す。
  - ①. 第一主成分の分散を最大化し、続く主成分はそれまでの主成分と直行

### 4. 2 問題

主成分分析とは、教師なし学習アルゴリズムの1つであり、**次元削減**を行うための手法である。

主成分分析は、**分散共分散行列**の**固有ベクトル**を基に新たな座標軸を複数作成し、新たな座標軸から構成される低次元空間へデータを写像する。この時、固有値の大きい固有ベクトルの方から第一主成分、第二主成分、・・・と呼ばれる。また、各主成分の固有値を固有値の総和で割った値が**寄与率**と呼ばれる。寄与率は各主成分がどの程度元のデータを説明できているのかを表す指標となる。

■分散共分散行列は半正定値行列であるため、固有値は0以上となる。そのため、負の固有値への対処は特に考える必要はない。

## 5. アルゴリズム

### 5. 1 K 近傍法 (KNN)

回帰と分類が行える手法。分類においては、特徴量空間において距離が近い順に任意の  $k$  個を取得し、多数決でデータが属するクラスを推定する

入力データに近い方から  $k$  個の学習データ (分類ラベル付き) を取得し、多数のものを取って、分類結果とする。

- ・柔軟にモデルを作れる。
- ・データの量が少ないと結果を発揮しにくい。

#### 1) 問題

入力されたデータに対し、すべての学習用データとの距離を計算し、最も近い学習用データのクラスに割り当てる分類アルゴリズムを **最近傍法** と呼ぶ。しかし、割り当てるクラスの決定に用いられるデータが 1 つのみでは、ノイズの影響を大きく受けてしまう。そこで、入力データに近い  $k$  個の学習データによる多数決によってクラスを割り当てる方法を **k 近傍法** と呼ぶ。

## 5. 1 K 平均法 (K-means)

データ点の所属するクラスを、各データ点からクラス重心への距離が最も近いものから選択する。

- ・手法が理解しやすく、大規模なデータにも適用可能。

### 1) 問題

教師なし学習とは、教師データを用いずに入力データ間の距離や類似度、統計的な性質に基づいてデータの構造を獲得するアルゴリズムを指す。

教師なし学習の代表的なタスクとして、入力データの類似度を基にグループ分けを行うクラスタリングがある。クラスタリングの代表的な手法として、グループを代表するベクトルの類似度に基づくグループ割り当てと、新たに割り当てられたグループに属するデータから、代表ベクトルの更新を繰り返す **k-means** がある。

しかし、**k-means** によって最終的に得られるグループ分けの結果は、代表ベクトルの初期値に依存するため、初期値を変えて複数回実行するなどの工夫が必要である。

**k-means** では各データが必ず 1 つのグループに分類されるがグループを代表する複数の正規分布を混合した **混合ガウスモデル** を用いることで確率的なクラスタリングを実現できる。**k-means** はデータの生成過程を確率モデルとして表現する手法であり、**生成モデル** の 1 つである。

## 6. サポートベクターマシン

### 6. 1 解説

識別境界近傍に位置する学習データ（サポートベクトル）と識別境界との距離であるマージンを最大化するように線形の境界を構築し、2クラス分類を行う。高度な数学的理論に支えられた手法。

明らかに所属クラスが分かる観測ではなく、判別境界の付近にある判断の難しい観測に着目する分類モデル。入力に用いる各データ点との距離が最大となるような境界線を求めることで、パターン分類を行う。

#### 1) マージンの最大化

- ・扱うデータは高次元
- ・データが線形分類できない（直線で分類できない）

#### 2) カーネル関数

- ・データをあえて高次元に写像することで、その写像後の空間で線形分類できるようにする。
- ・SVM やガウス過程などで利用される。
- ・有名なカーネル関数  
線形カーネル、RBF カーネル、多項式カーネル、周期的カーネル
- ・多層のニューラルネットワークに対応するカーネル関数

#### 3) カーネルトリック

写像に伴う計算量の増加を低く抑えるためのテクニック

### 6. 2 問題

1) サポートベクターマシン（SVM）は、主に2クラス分類に用いられるアルゴリズムである。

SVM は、ロジスティック回帰とは異なり、データとのマージンが最大となる識別面を学習によって獲得する。この時識別面を構成するために使われるデータをサポートベクトルと呼ぶ。マージン最大化は、ある学習データについて不等式制約のもとでの最適化問題を解くことで得られるが、その学習データが線形分離不可能な場合、全ての制約を満たす解を得ることができない。そこでスラック変数を導入することで、誤りをある程度許容しつつ、マージンを大きく取った識別面を得ることができる。このようなスラック変数を導入した SVM はソフトマージン SVM と呼ばれる

■SVM 最大の特徴は、マージン最大化を目的とした学習とスラック変数の導入である。