1. **線形回帰モデル**

**線形回帰とは、説明変数に対して目的変数が線形またはそれから近い値で表される状態のことで、統計学における回帰分析の一種であり、非線形回帰と対比される。 また線形回帰のうち、説明変数が1つの場合を単純線形回帰、2つ以上の場合を重回帰と呼ばれる(ウィキペディアより)。**

**線形回帰モデルは、目的変数と説明変数およびノイズ項(確率的誤差) の関係を以下のようにモデル化したものである。**

**モデルに含まれている予測子変数が1つだけである場合、そのモデルは単純な線形回帰モデルと呼ばれ、下記の式で表す。**

**ベクトル・行列記法を用いれば、線形回帰モデルは下記のように表せる。**

**機械学習では、を下記の表記がよく使われる。**

**ここで、 は 特徴量に対する「重み」をは「バイアス」を表し、直線の数式として考えると、それぞれ「傾き」と「切片」に該当する(Qiitaより)。**

**「重み」はパラメータと呼ばれ、未知の値で「最小二乗法」により推定する。**

**最小二乗法は、測定で得られた数値の組を、適当なモデルから想定される1次関数、対数曲線など特定の関数を用いて近似するときに、想定する関数が測定値に対してよい近似となるように、残差の二乗和を最小とするような係数を決定する方法、あるいはそのような方法によって近似を行うことである(ウィキペディアより)。**

**最小二乗法の数式は、モデル関数をとするとき下記数式を最小となるようにを求める式となる。**

1. **非線形回帰モデル**

**非線形回帰は、観測から得られたデータがモデルパラメータの非線形結合であり、1つ以上の独立した変数に依存する関数によってモデル化される回帰分析の一形式である。データは逐次近似法によって当て嵌められる(ウィキペディアより)。**

1. **基底展開法**

**基底関数とは、関数空間の基底ベクトルのことである。すなわち対象となる空間に属する全ての元（関数）は、この基底関数の線型結合で表される。**

**線形基底展開とは、を基底関数として、下記の形で展開する事。**

**例えば、実数値関数のフーリエ変換（コサイン変換・サイン変換）ではコサイン関数もしくはサイン関数、ウェーブレット変換ではウェーブレット関数とスケーリング関数、スプライン曲線では区分的多項式が基底関数として用いられる(ウィキペディアより)。**

**未知のパラメータは線形回帰モデルと同様に最小2乗法や最尤法により推定する。**

**線形回帰モデルは、説明変数に対して目的変数が線形なので、データの非線形性に対応できない。そこで、非線型性に対応する方法が基底展開法で、基底関数によって説明変数を非線形変換してから、線形回帰モデルで学習を行う。**

**このように、非線形モデルによって回帰タスクを解くことを、非線形回帰と呼ぶ(Qiitaより)。**

**基底展開法に基づく非線形回帰モデルは下記式で表される。**

**を基底関数という。**

1. **よく使われる基底関数**

* **多項式関数**
* **ガウス型基底関数**
* **スプライン関数/Bスプライン関数**

**曲線上の点の各座標値が，パラメータに関してスプライン関数で表される曲線。スプライン関数とは，多項式 (または有理式) をパラメータの区切り (ノット) において滑らかに接続した関数のことである(コトバンクより)。**

1. **ロジスティック回帰モデル**

**ロジスティック回帰は、ベルヌーイ分布に従う変数の統計的回帰モデルの一種である。連結関数としてロジットを使用する一般化線形モデル (GLM) の一種でもある。1958年にデイヴィッド・コックス（英語版）が発表した。確率の回帰であり、統計学の分類に主に使われる。医学や社会科学でもよく使われる。モデルは同じく1958年に発表された単純パーセプトロンと等価であるが、scikit-learnなどでは、パラメータを決める最適化問題で確率的勾配降下法を使用する物をパーセプトロンと呼び、座標降下法や準ニュートン法などを使用する物をロジスティック回帰と呼んでいる。**

**ロジスティック回帰モデルは、説明変数の情報に基づいて、データがどのクラスに属するのか、を分類したり、また、出来事が発生する確率を予測したりするために利用されるモデルである(Qiitaより)。**

**ロジスティック回帰では、対数オッズを説明変数 の線形和で表現する。予測したいこと（正事象）の確率をとしたとき、オッズは と書くことができ、正事象の起こりやすさを表します。オッズの対数をとったものが、対数オッズである。**

**ここで、重み は として切片を表します。ロジスティック回帰は、対数オッズと複数の説明変数の関係を表すモデルの重み を学習することが目的です。ただ、ロジスティック回帰を利用するときに関心があるのは、説明変数の値を与えたときの正事象の確率です。そこで、上式を左辺がになるように変形すると、**

**となる。モデルの重み を学習後、この式を利用して説明変数の値が与えられたときの正事象の確率を求めることができる。**

1. **主成分分析（PCA）**

**主成分分析は、相関のある多数の変数から相関のない少数で全体のばらつきを最もよく表す主成分と呼ばれる変数を合成する多変量解析の一手法。データの次元を削減するために用いられる（ウィキペディアより）。**

**主な数式は下記で表す。**

* **学習データ**

**※：説明変数のインデックス、：データの次元数**

**例えば、3次元から2次元への写像は、下記数式で表す。**

* **平均（ベクトル）**

**※：データ数**

* **データ行列**
* **分散共分散行列**
* **線形変換後のベクトル**

**※*j*：射影軸のインデックス、*a*：係数ベクトル**

1. **ラグランジュ関数**

**座標とその時間微分の関数で、ある物理系の力学的特性を表し、その運動を規定する量。フランスの物理学者・数学者ラグランジュが導入した。この関数に基づく運動方程式をラグランジュ方程式といい、その内容はニュートンの運動方程式と同等である。ラグランジュ関数をLとすると、系の運動エネルギーTとポテンシャルエネルギーUによりL=T-Uと定義される。Lと初期条件が与えられると系の全運動が記述されるので、これから出発する力学の形式をラグランジュ形式とよび、解析力学の基本的な体系となっている（コトバンクより）。**

**ラグランジュ関数を用いて例として、以下の制約付き最適化問題を解く。制約を入れないと解が無限となるためノルムが1となる制約を入れる。**

**※ノルム は、平面あるいは空間における幾何学的ベクトルの "長さ" の概念の一般化であり、ベクトル空間に対して「距離」を与えるための数学の道具である（ウィキペディアより）。**

* **目的関数**
* **制約条件**
* **制約付き最適化問題の解き方**

**ラグランジュ関数を最大にする係数ベクトルを探索する。**

**※微分して0になる点を見つける。**

* **ラグランジュ関数**
* **ラグランジュ関数を微分して最適解を求める**

**となり、元のデータの分散共分散行列の固有値と固有ベクトルを求めればよいことが分かる。**

1. **寄与率（決定係数）**

**決定係数は、統計学において、独立変数（説明変数）が従属変数（目的変数）のどれくらいを説明できるかを表す値である。寄与率と呼ばれることもある。標本値から求めた回帰方程式（モデル）のあてはまりの良さの尺度として利用される（ウィキペディアより）。**

**機械学習での表現を下記に記す。**

* **第１~元次元分の主成分の分散は、元のデータの分散と一致**

**2次元のデータを2次元の主成分で表示した時、固有値の和と元のデータの分散が一致する。**

**第主成分の分散は主成分に対応する固有値である。**

* **寄与率**

**第主成分の分散の全分散に対する割合(第kk主成分が持つ情報量の割合)のこと。**

**：第主成分の分散**

**：主成分の総分数**

* **累積寄与率**

**累積寄与率とは，第1主成分から第主成分までの寄与率の和のことで、第1主成分から第主成分での圧縮がデータの散らばり具合をどの程度カバーしているかの説明する割合である。1に近ければデータの散らばり具合を説明できている割合が高いことになる（Qiitaより）。**

**：第1~主成分の分散**

**：主成分の総分数**

1. **アルゴリズム**

**機械学習で使用されるアルゴリズムには、次のアルゴリズムがある（Qiita）。**

**決定木、ランダムフォレスト、ロジスティック回帰、サポートベクターマシン（SVM）、ナイーブベイズ分類器、近傍法、平均法（）、アダブースト、ニューラルネットワーク、マルコフ連鎖。講義で学習した下記のアルゴリズムについて記述する。**

1. **近傍法**

**近傍法は、特徴空間における最も近い訓練例に基づいた分類の手法であり、パターン認識でよく使われる。近傍法は、機械学習アルゴリズムの中でも簡単なアルゴリズムと言われいている。理由は、インスタンスの分類を、その近傍のオブジェクト群の多数決で行うことで決定するからである（Qiitaより）。**

1. **平均法（k-means法）**

**平均法は、クラスタリングと呼ばれる、データを性質の近い分類同士でグループ分けするためのアルゴリズムのひとつです。クラスタリングの最も簡単な手法の一つであり，教師なし学習です（Qiitaより）。**

**平均法は下記工程で行われる。**

1. **各点に対してランダムにクラスタを割り振る。**
2. **各クラスタに割り当てられた点について重心を計算する。**
3. **各点について上記で計算された重心からの距離を計算し、距離が一番近いクラスタに割り当て直す。**
4. **2.と3.の工程を、割り当てられるクラスタが変化しなくなるまで行う。**
5. **サポートベクターマシーン（SVM）**

**サポートベクターマシンは、教師あり学習を用いるパターン認識モデルの一つで、線形入力素子を利用して2クラスのパターン識別器を構成する手法である（Qiitaより）。**

1. **ハードマージンとソフトマージン**

**線形分離可能(一つの直線で二つに分けられる)なデータを前提としたマージンをハードマージンと呼び、線形分離不可能なデータを前提として、誤判別を許容するマージンをソフトマージンと呼ぶ。**

**次元のデータを分類する次元の平面(厳密には平面ではない)を分離超平面と呼ぶ。また分離超平面とその分離超平面に最も近いデータとの距離をマージンと呼び、このマージンを最大化することがこのアルゴリズムの目標になる。分離超平面に最も近いでデータのことをサポートベクトルと呼ぶ（Qiitaより）。**

1. **次元数ベクトル空間の超平面の式**

**全部で個のデータが存在する場合、次元数ベクトル空間上の超平面の数式を下記とする。**

1. **ハードマージン最適化の式を導出**

**を計算すると、となり、二次元における直線の式であるは、次元に拡張した超平面の式であることが推測できる。**

**２つのクラスに分けられる集合をとすると、以下の式が導出される。**

**この式をまとめて表すためにラベル変数を導入する。**

**番目のデータがクラス1に属するときに、クラス2に属するときに**

**とすると、下記式となる。**

**このように定義したを用いると、条件式を以下のように表すことができる**