1　序論

ここは読み合わせとする。要点は下記の通り。

・データ分析と物理学の歴史*1*2

・機械学習の基本的な概念や学習方法の種類*3

1.1　多項式曲線フィッティング

ここでの目的は sklearn を用いて多項式曲線フィッティングの実装を体験することである。教科書の例と同様に多項式回帰による sin 関数の実装・及び汎化性能の評価を行う。(→ jupyter へ)

1.2　確率論

ここは読み合わせとする。要点は下記の通り。

・確率の加法性*4

・条件付確率の定義とベイズの定理・事前確率と事後確率の定義

1.2.1　確率密度

基本的には読み合わせ。要点は下記の通り。

連続変数に関する

・確率密度関数 (p.d.f)*5の定義

・累積分布関数 (c.d.f)*6の定義

なお、本レジュメにおいては、p.d.f. の mode と変数変換について議論を行う。特に、確率変数 (r.v.)*7に非線形な変数変換を施すと mode が変化してしまうことを説明する。なお、編集の都合上、ここの説明部分は次頁以降に手書きでまとめているので留意されたい。

1.2.2　期待値と分散

ここは読み合わせとする。要点は下記の通り。

・期待値と分散の定義

・PRML 内における上記の表現方法

*1 ティコ・ブラーエ (1546-1601) は六分儀、象限儀と呼ばれる観測器具を用いて恒星と惑星の位置を観測し、膨大なデータを収集しました。その観測精度は約 2 分の誤差であり従来の 5 倍ほど高い精度を有していたそうです。望遠鏡を使わずして、この精度を得たことは素晴らしい功績でした。彼は病で 54 年のその生涯を終えました。

*2 ヨハネス・ケプラー (1571-1630) が解析したのは天体の動きを中心力場における 2 体問題に落とし込んだものです。運動方程式の積分により、質点の軌跡の方程式を求めました。ちなみに、先述したティコ・ブラーエは「正確な経験的事実を追い求める人物」と評されていました。実際その経験的事実については弟子のケプラーが、師の集めてくれたデータをもとに導いてくれたのでした。ただ、ティコ・ブラーエが早死にしたのは、ケプラーが毒を盛ったせいなのではないかという説も囁かれています。

*3 釈迦に説法な気がするので読み合わせさえもスキップしたいくらいです。

*4 コルモゴロフの公理の 1 つでした。念のために書いておくと「各事象 A_i が互いに排反であるとき、$P(U_{i=1}^{\infty}A_i) = \Sigma_{i=1}^{\infty}P(A_i)$ である」というものでした。

*5 Probability density function の略

*6 Cumulative distribution function の略

*7 Random value の略

1.2.1

Modeと変数変換について以下2点をクリアにする.

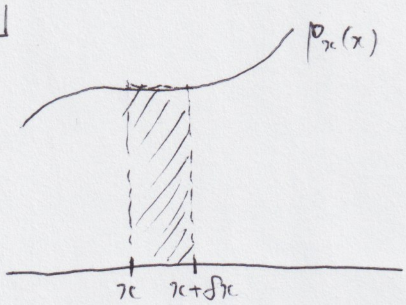
① (1.27)式の絶対値が付く理由

② (非線形)変数変換によるmodeの変更理由

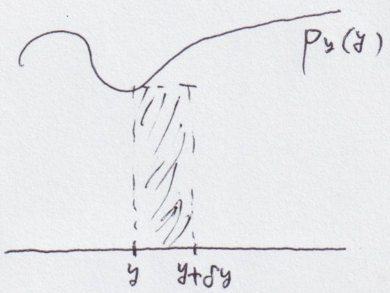
①についてヴィジュアルで理解したい. $x=g(y)$という変数変換を考える. このとき区間$(x, x+\delta x)$, $(y, y+\delta y)$とp.d.f. $p_x(x)$, $p_y(y)$を図示する。

Case.1

x」



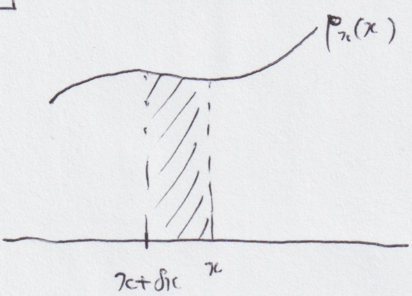
y」

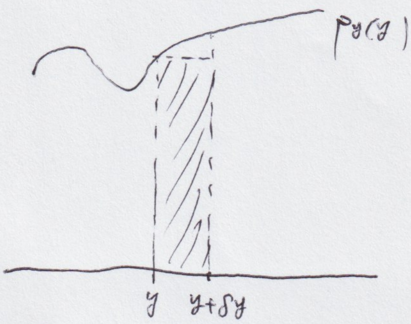


上では$\delta x, \delta y > 0$という仮定のもとで図示したが、次のような場合もありえる.

Case.2

x」





Case.1において. 各区間に事象が入る確率は

$x \sim x+\delta x$; $p(x)\delta x$

$y \sim y+\delta y$; $p(y)\delta y$

と表現され. δxやδyが十分小さければ

$p(x)\delta x \fallingdotseq p(y)\delta y$

とすることができる. このときはδをdに置き換えることで

$$p_y(y) = p_x(x)\frac{dx}{dy}$$

が導かれる.

しかし、Case 2 において、各区間に事象が入る確率は

$x+\delta x \sim x$ ： $p(x)(-\delta x)$ （$\because \delta x < 0$）

$y \sim y+\delta y$ ： $p(y)\delta(y)$

と表現されるため、両者をイコールでつないだ後、負号を消すため絶対値を付けることになる。

$$p(x)(-\delta x) = p(y)\delta(y)$$

$$|p(x)\delta x| = |p(y)\delta(y)|$$

$$p(x)|\delta x| = p(y)|\delta(y)|$$

$$\longrightarrow p_y(y) = p_x(x)\left|\frac{dx}{dy}\right|$$

ただし、これまでの議論において $p_x(x), p_y(y) \geq 0$ は暗に仮定している。

実際、$\delta x, \delta y$ の符号がそろっていても絶対値を付けることには何の問題もない。

したがい、(1.27)式では絶対値を用いた表現をしている。

②について、演習1.4に取り組むこととしたい。示すこととしては、「x, y それぞれの mode を与える $\hat{x}, \hat{y}$ が変数変換の関係 $\hat{x} = g(\hat{y})$ でないこと」である。

ある y について $\dfrac{dx}{dy} \geq 0$ のときを考える。(1.27)は

$$p_y(y) = p_x(x)\frac{dx}{dy}$$

となる。ここで $p_y(y)$ を y について max 化する必要条件は

$$\frac{dp_y(y)}{dy} = 0$$

である。つまり

$$\frac{dp_y(y)}{dy} = \underbrace{\frac{dp_x(x)}{dx}\cdot\frac{dx}{dy}}_{\text{合成関数の微分}}\cdot\frac{dx}{dy} + p_x(x)\frac{d^2x}{dy^2} = 0 \quad \cdots (*)$$

ここで $\hat{x}$ が $p_x(x)$ の max を与えるとすると $\left.\dfrac{dp_x(x)}{dx}\right|_{x=\hat{x}} = 0$ となる。したがい

(*) の第2辺の第1項へ $x=\hat{x}$ を代入すると、この第1項は0になる。しかし、第2項については $\dfrac{d^2g(y)}{dy^2}$ という部分が存在し、g が y の非線形関数である場合は0にならない。

すなわち(*)の2つめのイコールは成立しない。したがって、$\hat{x}$が$p_x(x)$のmaxを与えるとき、この$\hat{x}$を使って$\hat{x} = g(\tilde{y})$をみたす$\tilde{y}$を作っても

$$\left.\frac{dp_y(y)}{dy}\right|_{y=\tilde{y}} \neq 0$$ となってしまい $$\left.\frac{dp_y(y)}{dy}\right|_{y=\hat{y}} = 0$$ となる$\hat{y}$とは

異なるyの値を得てしまう。よって一般的には$\hat{x} \neq g(\hat{y})$である。

ただし、$x = g(y) = ay + b$、つまり線形変換の場合は

$$\frac{d^2x}{dy^2} = \frac{d^2}{dy^2}(ay+b) = 0$$

となるため、$\left.\frac{dp_x(x)}{dx}\right|_{x=\hat{x}} = 0$のとき、(*)から$\left.\frac{dp_y(y)}{dy}\right|_{y=\hat{y}} = 0$である。

つまりgが線形ならば$\hat{x} = g(\hat{y})$となる。

$\frac{dx}{dy} < 0$でも同様である。