

1 近似推論法

EM アルゴリズムが困難な場合

- EM アルゴリズムで行う計算

- **E ステップ**では、潜在変数の事後確率分布 $p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}})$ を計算
- **M ステップ**では、完全データ対数尤度 $\ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta})$ の期待値を計算

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = \sum_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}}) \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}) \quad (1)$$

$$Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) = \int_{\mathbf{Z}} p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}}) \ln p(\mathbf{X}, \mathbf{Z}|\boldsymbol{\theta}) d\mathbf{Z} \quad (2)$$

そして、 $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$ を最大化するパラメータ $\boldsymbol{\theta}^{\text{new}}$ を求める

$$\boldsymbol{\theta}^{\text{new}} = \arg \max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}}) \quad (3)$$

EM アルゴリズムが困難な場合

● EM アルゴリズムの困難さ

- 実際に扱うモデルでは、事後分布 $p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta}^{\text{old}})$ の計算や、事後分布に従った期待値 $Q(\boldsymbol{\theta}, \boldsymbol{\theta}^{\text{old}})$ の計算が、**不可能であることが多い**
- 隠れ変数の**次元が高すぎる**
- 事後分布が**複雑な形をしていて**、期待値を解析的に計算できない
- **連続変数**であれば、積分が閉形式の解を持たないかもしれない
- 空間の次元の問題や、被積分項の複雑さから、数値積分すら困難かもしれない
- **離散変数**であれば、期待値を計算するためには、**潜在変数の可能な全ての組み合わせについての和を取る**必要がある
- 隠れ変数の次元が高くなると、組み合わせ数が指数的に増大する
- 計算量が大きすぎて、期待値の厳密な計算がもはや不可能

EM アルゴリズムが困難な場合

- 近似法

- EM アルゴリズムが困難であるとき、何らかの方法で**近似**しなければならない
- 近似法は、**確率的な近似**と、**決定的な近似**の 2 つに分けられる

- 確率的な近似

- **マルコフ連鎖モンテカルロ法**などの手法がある
- **無限の計算資源があれば**、厳密な結果が得られる
- 実際には計算量が有限であるため、得られる解は近似解となる

- 決定的な近似

- 事後分布 $p(\mathbf{Z}|\mathbf{X}, \boldsymbol{\theta})$ を**解析的に近似**する
- 事後分布に対して、**何らかの仮定**をおく
- 例えば、**単純な項の積として分解できる**、あるいは、(ガウス分布などの特別な) **パラメトリックな分布である**といった仮定

- ここで扱う近似法
 - 変分推論法 (Variational inference) あるいは変分ベイズ法 (Variational Bayes) について扱う
- 変分推論 (Variational inference)
 - 18 世紀のオイラー、ラグランジュらによる変分法 (Calculus of variations) に起源をもつ
 - まずは、変分法について説明をしていく

- 関数と汎関数の違い

- 通常関数は、入力として値をとり、出力として関数の値を返す
- 通常関数は、値から値への写像である
- 関数の導関数は、入力値を微小に変えたときに、出力の関数値がどの程度変わるかを表す
- 汎関数 (Functional) とは、入力として関数を取り、出力として汎関数の値を返す
- 汎関数は、関数から値への写像である
- 汎関数微分 (Functional derivative) とは、入力関数が微小に変わったときに、出力の汎関数値がどの程度変わるかを表す
- 汎関数の微分を、変分という

- 汎関数の例

- エントロピー $H[p]$ は、確率分布 $p(x)$ を入力として、以下の量を返す汎関数

$$H[p] = - \int p(x) \ln p(x) dx \quad (4)$$

- 汎関数の最適化

- 多くの問題は、**汎関数の値を最適化する問題**として定式化できる
- 汎関数の最適化とは、**可能な全ての入力関数の中から**、汎関数の値を最大化、あるいは最小化するような**関数を選び出す**ことである
- 通常最適化では、可能な全てのパラメータ (入力値) の中から、関数を最大化、あるいは最小化するような 1 つのパラメータを選び出す
- 次は、いよいよ**変分**の計算について説明する

- 変分法

- 通常の微分を使えば、ある関数 $y(x)$ を最大化 (最小化) するような x の値が求められる
- **変分法**を使えば、汎関数 $F[y]$ を最大化 (最小化) するような、関数 $y(x)$ が求められる
- 従って、可能な全ての関数 $y(x)$ の中から、 $F[y]$ を最大 (最小) にするような関数が得られる

- 変分法によって解ける問題の例

- 2 点を結ぶ最短経路は? (答えは直線)
- 最速降下曲線は? (答えはサイクロイド)
- **エントロピーが最大**になるような確率分布は? (答えは**ガウス分布**)

- 通常の微分の表現

- 関数 $y(x + \epsilon)$ のテイラー展開は次のように記述できた

$$y(x + \epsilon) = \sum_{n=0}^{\infty} \frac{y^{(n)}(x)}{n!} \epsilon^n \quad (5)$$

$$= y(x) + \frac{dy}{dx} \epsilon + \frac{1}{2!} \frac{d^2 y}{dx^2} \epsilon^2 + \frac{1}{3!} \frac{d^3 y}{dx^3} \epsilon^3 + \dots \quad (6)$$

$$= y(x) + \frac{dy}{dx} \epsilon + O(\epsilon^2) \quad (7)$$

- これより微分 dy/dx は、次のように求められる
- 変数 x に微小な変化 ϵ を加え、このときの関数値 $y(x + \epsilon)$ を ϵ の累乗形として表現する
- 最後に $\epsilon \rightarrow 0$ の極限をとればよい

$$\frac{dy}{dx} = \lim_{\epsilon \rightarrow 0} \frac{y(x + \epsilon) - y(x)}{\epsilon} \quad (8)$$

- 多変数関数 $y(x_1, \dots, x_D)$ の偏微分の表現
 - 多変数関数 $y(x_1, \dots, x_D)$ のテイラー展開は次のように記述できた

$$\mathbf{D}^n = \left(\epsilon_1 \frac{\partial y}{\partial x_1} + \dots + \epsilon_D \frac{\partial y}{\partial x_D} \right)^n \quad (9)$$

上記のような演算子 \mathbf{D} を考えれば

$$\begin{aligned} & y(x_1 + \epsilon_1, \dots, x_D + \epsilon_D) \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} (\mathbf{D}^n y)(x_1, \dots, x_D) \end{aligned} \quad (10)$$

$$\begin{aligned} &= y(x_1, \dots, x_D) + \sum_{i=1}^D \frac{\partial y}{\partial x_i} \epsilon_i + \frac{1}{2!} \sum_{i=1}^D \sum_{j=1}^D \frac{\partial^2 y}{\partial x_i \partial x_j} \epsilon_i \epsilon_j + \\ & \quad \frac{1}{3!} \sum_{i=1}^D \sum_{j=1}^D \sum_{k=1}^D \frac{\partial^3 y}{\partial x_i \partial x_j \partial x_k} \epsilon_i \epsilon_j \epsilon_k + \dots \end{aligned} \quad (11)$$

であるから

$$\begin{aligned} & y(x_1 + \epsilon_1, \dots, x_D + \epsilon_D) \\ = & y(x_1, \dots, x_D) + \sum_{i=1}^D \frac{\partial y}{\partial x_i} \epsilon_i + O(\epsilon^2) \end{aligned} \quad (12)$$

- これより偏微分 $\partial y / \partial x_i$ は、次のように求められる

$$\begin{aligned} \frac{\partial y}{\partial x_i} = \lim_{\epsilon_i \rightarrow 0} \frac{1}{\epsilon_i} & (y(x_1, \dots, x_{i-1}, x_i + \epsilon_i, x_{i+1}, \dots, x_D) - \\ & y(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_D)) \end{aligned} \quad (13)$$

- 変分の表現

- 多少不正確だが、変分をどのように定義すればよいか考えてみる
- ここで、各 x_i に対する関数の値 $z_i = y(x_i)$ を個別の変数とみなして、次の関数 $F(z_1, \dots, z_D)$ について考えてみよう

$$\begin{aligned} & F(z_1 + \epsilon\eta(x_1), \dots, z_D + \epsilon\eta(x_D)) \\ &= F(z_1, \dots, z_D) + \sum_{i=1}^D \frac{\partial F}{\partial z_i} \epsilon\eta(x_i) + O(\epsilon^2) \end{aligned} \quad (14)$$

$z_i = y(x_i)$ を代入してみると

$$\begin{aligned} & F(y(x_1) + \epsilon\eta(x_1), \dots, y(x_D) + \epsilon\eta(x_D)) \\ &= F(y(x_1), \dots, y(x_D)) + \sum_{i=1}^D \frac{\partial F}{\partial y(x_i)} \epsilon\eta(x_i) + O(\epsilon^2) \end{aligned} \quad (15)$$

変分法

- ここで $D \rightarrow \infty$ の極限を取り、 x_1, \dots, x_D が、ある連続した区間 $[a, b]$ に含まれる、全ての実数を表すことにする
- このとき $y(x_1), \dots, y(x_D)$ は、実数の区間 $[a, b]$ で定義される連続関数 $y(x)$ として書けることが分かる
- 同様に $y(x_1) + \epsilon\eta(x_1), \dots, y(x_D) + \epsilon\eta(x_D)$ は、実数の区間 $[a, b]$ で定義される連続関数 $y(x) + \epsilon\eta(x)$ として、まとめることができる
- 関数 $\eta(x)$ も、実数の区間 $[a, b]$ で定義される連続関数
- $\epsilon\eta(x)$ は、 $y(x)$ に加わる摂動として、考えることができる

変分法

- 関数 F は、関数 $y(x)$ や $y(x) + \epsilon\eta(x)$ を入力として受け取る、汎関数 $F[y]$ として解釈できるから、次のように書ける

$$F(y(x_1) + \epsilon\eta(x_1), \dots, y(x_D) + \epsilon\eta(x_D)) = F[y(x) + \epsilon\eta(x)] \quad (16)$$

$$F(y(x_1), \dots, y(x_D)) = F[y(x)] \quad (17)$$

- 以下の項は、入力を $y(x)$ に摂動を加えて $y(x) + \epsilon\eta(x)$ へと微小に変化させたときの、汎関数の ($F[y(x)]$ から $F[y(x) + \epsilon\eta(x)]$ への) 変化量を表している

$$\sum_{i=1}^D \frac{\partial F}{\partial y(x_i)} \epsilon\eta(x_i) \quad (18)$$

- 点 x_i における汎関数 F の変化量を、 x_1, \dots, x_D の範囲について、即ち、実数の区間 $[a, b]$ について足し合わせていると解釈する

変分法

- $D \rightarrow \infty$ のとき、 x_1, \dots, x_D は区間 $[a, b]$ における全ての実数を表すから、総和を積分に置き換えられそうである
- 汎関数の微分 $\frac{\delta F}{\delta y(x)}$ を使えば、次のように書ける

$$\begin{aligned} & \sum_{i=1}^D \frac{\partial F}{\partial y(x_i)} \epsilon \eta(x_i) \\ \Rightarrow & \int_a^b \frac{\delta F}{\delta y(x)} \epsilon \eta(x) dx = \epsilon \int_a^b \frac{\delta F}{\delta y(x)} \eta(x) dx \end{aligned} \quad (19)$$

- 結局、変分 $\frac{\delta F}{\delta y(x)}$ は次のように定義できる

$$F[y(x) + \epsilon \eta(x)] = F[y(x)] + \epsilon \int_a^b \frac{\delta F}{\delta y(x)} \eta(x) dx + O(\epsilon^2) \quad (20)$$

変分法

- $F[y]$ は、区間 $[a, b]$ で定義される関数 y を受け取るとする
- 変分 $\delta F/\delta y$ は、入力関数 $y(x)$ に、任意の微小な変化 $\epsilon\eta(x)$ を加えたときの、汎関数 $F[y]$ の変化量として定義できる
- $\eta(x)$ は x についての任意の関数

Figure D.1 A functional derivative can be defined by considering how the value of a functional $F[y]$ changes when the function $y(x)$ is changed to $y(x) + \epsilon\eta(x)$ where $\eta(x)$ is an arbitrary function of x .

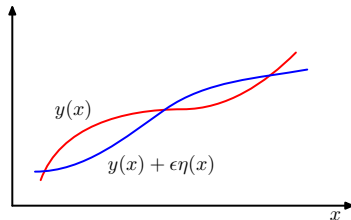


図 1: $y(x)$ と $y(x) + \epsilon\eta(x)$ の表現

- 変分法の例

- 次の図 2 を使って、実際に変分を求めてみよう

- 汎関数 $F[y]$ は、次のように定義されとする

$$F[y] = \int_a^b y(x) dx \quad (21)$$

- 汎関数の値 $F[y(x)], F[y(x) + \epsilon\eta(x)]$ は次のようになる

$$F[y(x)] = \int_a^b y(x) dx \quad (22)$$

$$F[y(x) + \epsilon\eta(x)] = \int_a^b (y(x) + \epsilon\eta(x)) dx \quad (23)$$

- $F[y(x) + \epsilon\eta(x)]$ は次のように分解できる

$$F[y(x) + \epsilon\eta(x)] = \int_a^b y(x)dx + \epsilon \int_a^b \eta(x)dx \quad (24)$$

$$= F[y(x)] + \epsilon \int_a^b \eta(x)dx \quad (25)$$

- ここで、変分の定義式は

$$F[y(x) + \epsilon\eta(x)] = F[y(x)] + \epsilon \int_a^b \frac{\delta F}{\delta y(x)} \eta(x)dx + O(\epsilon^2) \quad (26)$$

であったので、上の2つの式を見比べれば、変分 $\delta F/\delta y$ は結局

$$\frac{\delta F}{\delta y(x)} = 1 \quad (27)$$

となることが分かる

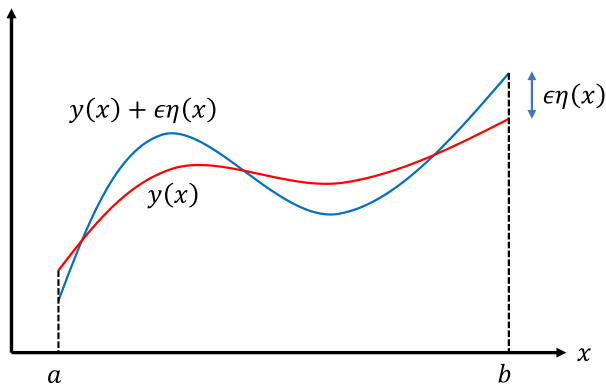


図 2: 区間 $[a, b]$ で定義された関数 $y(x)$ の表現

● 汎関数の最適化

- 汎関数 $F[y]$ が最大 (最小) となるとき、関数 $y(x)$ の微小な変化に対して、汎関数は変化しないはず
- 即ち、汎関数が最大 (最小) となるとき、 $F[y(x) + \epsilon\eta(x)] = F[y(x)]$ が成り立つ
- 従って、変分の定義式から、以下が成り立つ

$$\int_a^b \frac{\delta F}{\delta y(x)} \eta(x) dx = 0 \quad (28)$$

- 上式は任意の $\eta(x)$ について成立しなければならない
- 従って、変分 $\delta F/\delta y$ は、任意の x について 0 とならなければならない
- 汎関数 $F[y]$ が最大 (最小) となるとき、 $\delta F/\delta y = 0$ が成立することが分かった (通常の微分と同じ)

- 変分法の例

- 様々な汎関数について、変分を導出してみよう
- また、その汎関数が最大 (最小) となるときに成り立つ条件を、導出してみよう

- 汎関数の例 (1)

- $y(x)$ とその微分 $y'(x) = dy/dx$ 、そして x によって決まる関数 $G(y(x), y'(x), x)$ があるとする
- 汎関数 $F[y]$ を、 $G(y(x), y'(x), x)$ を区間 $[a, b]$ にわたって積分した結果を出力する関数として、次のように定める

$$F[y] = \int_a^b G(y(x), y'(x), x) dx \quad (29)$$

- 積分区間は無限であってもよいとする ($a = -\infty, b = \infty$ でもよい)

変分法

- $y(x)$ に摂動 $\epsilon\eta(x)$ を加えたときの、汎関数の値 $F[y(x) + \epsilon\eta(x)]$ を使って、変分 $\delta F/\delta y$ を調べてみる

$$F[y(x) + \epsilon\eta(x)] = \int_a^b G(y(x) + \epsilon\eta(x), y'(x) + \epsilon\eta'(x), x) dx \quad (30)$$

ここで、被積分項のテーラー展開を考えれば

$$\begin{aligned} & G(y(x) + \epsilon\eta(x), y'(x) + \epsilon\eta'(x), x) \\ = & G(y(x), y'(x), x) + \frac{\partial G}{\partial y} \epsilon\eta(x) + \\ & \frac{\partial G}{\partial y'} \epsilon\eta'(x) + \frac{\partial G}{\partial x} \cdot 0 + O(\epsilon^2) \end{aligned} \quad (31)$$

$$= G(y(x), y'(x), x) + \epsilon \left(\frac{\partial G}{\partial y} \eta(x) + \frac{\partial G}{\partial y'} \eta'(x) \right) + O(\epsilon^2) \quad (32)$$

であるから

$$\begin{aligned}
 & F[y(x) + \epsilon \eta(x)] \\
 = & \int_a^b G(y(x) + \epsilon \eta(x), y'(x) + \epsilon \eta'(x), x) dx \\
 = & \int_a^b \left(G(y(x), y'(x), x) + \right. \\
 & \left. \epsilon \left(\frac{\partial G}{\partial y} \eta(x) + \frac{\partial G}{\partial y'} \eta'(x) \right) + O(\epsilon^2) \right) dx \quad (33)
 \end{aligned}$$

$$\begin{aligned}
 = & \int_a^b G(y(x), y'(x), x) dx + \\
 & \epsilon \int_a^b \left(\frac{\partial G}{\partial y} \eta(x) + \frac{\partial G}{\partial y'} \eta'(x) \right) dx + O(\epsilon^2) \quad (34)
 \end{aligned}$$

$$= F[y(x)] + \epsilon \int_a^b \left(\frac{\partial G}{\partial y} \eta(x) + \frac{\partial G}{\partial y'} \eta'(x) \right) dx + O(\epsilon^2) \quad (35)$$

ここで

$$\begin{aligned} & \int_a^b \left(\frac{\partial G}{\partial y} \eta(x) + \frac{\partial G}{\partial y'} \eta'(x) \right) dx \\ = & \int_a^b \left(\frac{\partial G}{\partial y} \eta(x) \right) dx + \int_a^b \left(\frac{\partial G}{\partial y'} \eta'(x) \right) dx \end{aligned} \quad (36)$$

$$\begin{aligned} = & \int_a^b \left(\frac{\partial G}{\partial y} \eta(x) \right) dx + \\ & \left[\frac{\partial G}{\partial y'} \eta(x) \right]_a^b - \int_a^b \frac{d}{dx} \left(\frac{\partial G}{\partial y'} \right) \eta(x) dx \end{aligned} \quad (37)$$

$$= \left[\frac{\partial G}{\partial y'} \eta(x) \right]_a^b + \int_a^b \left(\frac{\partial G}{\partial y} - \frac{d}{dx} \left(\frac{\partial G}{\partial y'} \right) \right) \eta(x) dx \quad (38)$$

である

変分法

- 途中の式変形では、部分積分を使っていることに注意
- いま、積分区間の両端において、 $y(x)$ の値は固定されているとする
- これを **固定端条件** という (図 3)
- このとき、 $\eta(a) = \eta(b) = 0$ であるから、上式の最初の項が消えて

$$\begin{aligned} & \int_a^b \left(\frac{\partial G}{\partial y} \eta(x) + \frac{\partial G}{\partial y'} \eta'(x) \right) dx \\ &= \left[\frac{\partial G}{\partial y'} \eta(x) \right]_a^b + \int_a^b \left(\frac{\partial G}{\partial y} - \frac{d}{dx} \left(\frac{\partial G}{\partial y'} \right) \right) \eta(x) dx \end{aligned} \quad (39)$$

$$= \int_a^b \left(\frac{\partial G}{\partial y} - \frac{d}{dx} \left(\frac{\partial G}{\partial y'} \right) \right) \eta(x) dx \quad (40)$$

のようになる

変分法

- 従って、摂動を加えたときの汎関数の値 $F[y(x) + \epsilon\eta(x)]$ は

$$\begin{aligned} & F[y(x) + \epsilon\eta(x)] \\ = & F[y(x)] + \epsilon \int_a^b \left(\frac{\partial G}{\partial y} - \frac{d}{dx} \left(\frac{\partial G}{\partial y'} \right) \right) \eta(x) dx + O(\epsilon^2) \quad (41) \end{aligned}$$

となる

- 上式を、変分の定義式と比べれば

$$F[y(x) + \epsilon\eta(x)] = F[y(x)] + \epsilon \int_a^b \frac{\delta F}{\delta y(x)} \eta(x) dx + O(\epsilon^2) \quad (42)$$

変分 $\delta F/\delta y$ は次のように書ける

$$\frac{\delta F}{\delta y(x)} = \frac{\partial G}{\partial y} - \frac{d}{dx} \left(\frac{\partial G}{\partial y'} \right) \quad (43)$$

変分法

- 汎関数 $F[y]$ が最大 (最小) になるとき、変分 $\delta F/\delta y$ が 0 になる
- 従って、汎関数が最大 (最小) になるとき、以下の方程式が成り立つ

$$\frac{\partial G}{\partial y} - \frac{d}{dx} \left(\frac{\partial G}{\partial y'} \right) = 0 \quad (44)$$

- これをオイラー-ラグランジュ方程式という
- オイラー-ラグランジュ方程式は、次のような考え方で導出することもできる
- $F[y]$ が最大 (最小) であれば、摂動 $\epsilon\eta(x)$ によって $y(x)$ が少し変化しても、 $F[y]$ の値は変化しないはず
- 従って、 $F[y]$ が最大 (最小) であるとき、 $F[y]$ の ϵ による微分は 0 になるはず

- これを数式で表現すると、次のようになる

$$\left. \frac{\partial F[y]}{\partial \epsilon} \right|_{\epsilon=0} = 0 \quad (45)$$

左辺は通常の偏微分であり、これを計算すると

$$\begin{aligned} & \frac{\partial F[y]}{\partial \epsilon} \\ = & \frac{\partial}{\partial \epsilon} \int_a^b G(y, y', x) dx \end{aligned} \quad (46)$$

$$= \int_a^b \frac{\partial}{\partial \epsilon} G(y, y', x) dx \quad (47)$$

$$= \int_a^b \left(\frac{\partial G}{\partial y} \frac{\partial y}{\partial \epsilon} + \frac{\partial G}{\partial y'} \frac{\partial y'}{\partial \epsilon} + \frac{\partial G}{\partial x} \frac{\partial x}{\partial \epsilon} \right) dx \quad (48)$$

$$= \int_a^b \left(\frac{\partial G}{\partial y} \eta(x) + \frac{\partial G}{\partial y'} \eta'(x) \right) dx \quad (49)$$

$$= \left[\frac{\partial G}{\partial y'} \eta(x) \right]_a^b + \int_a^b \left(\frac{\partial G}{\partial y} - \frac{d}{dx} \left(\frac{\partial G}{\partial y'} \right) \right) \eta(x) dx \quad (50)$$

($\because \eta(a) = \eta(b) = 0$)

$$= \int_a^b \left(\frac{\partial G}{\partial y} - \frac{d}{dx} \left(\frac{\partial G}{\partial y'} \right) \right) \eta(x) dx \quad (51)$$

$$= 0$$

- 上の式変形では、 $y = y(x) + \epsilon \eta(x)$ であるから

$$\frac{\partial y}{\partial \epsilon} = \eta(x) \quad (52)$$

$$\frac{\partial y'}{\partial \epsilon} = \frac{\partial}{\partial \epsilon} \left(\frac{\partial y}{\partial x} \right) = \frac{\partial}{\partial \epsilon} (y'(x) + \epsilon \eta'(x)) = \eta'(x) \quad (53)$$

$$\frac{\partial x}{\partial \epsilon} = 0 \quad (54)$$

が成立することを利用している

- 任意の $\eta(x)$ について、上式が恒等的に成り立つためには

$$\frac{\partial G}{\partial y} - \frac{d}{dx} \left(\frac{\partial G}{\partial y'} \right) = 0 \quad (55)$$

でなければならないことが分かり、先程と同様に、オイラー-ラグランジュ方程式を得る

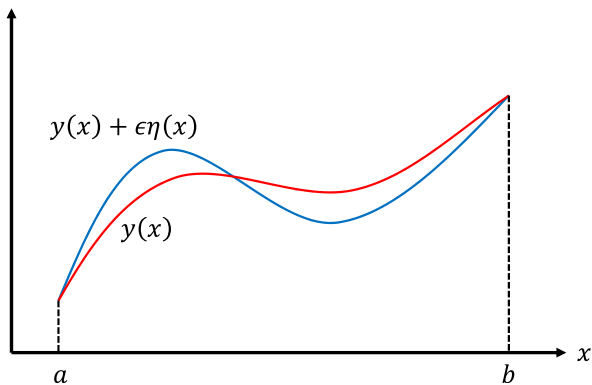


図 3: 制約条件を含んでいる場合の表現

- 汎関数の例 (2)

- 上では $G(y(x), y'(x), x)$ について考えて、変分を導出した
- $y(x)$ と x のみによって決まり、 $y'(x)$ には依存しない関数 $G(y(x), x)$ を考えよう
- 汎関数 $F[y]$ は、先程と同様に以下で表されとする

$$F[y] = \int_a^b G(y(x), x) dx \quad (56)$$

- このとき変分 $\delta F / \delta y$ を求めるのは、非常に簡単である
- 先程の式に、 $\partial G / \partial y' = 0$ を代入すれば直ちに得られる

$$\frac{\delta F}{\delta y(x)} = \frac{\partial G}{\partial y} \quad (57)$$

- あるいは以下のように書ける

$$\frac{\delta}{\delta y(x)} \int_a^b G(y(x), x) dx = \frac{\partial}{\partial y} G(y(x), x) \quad (58)$$

- $F[y]$ が最大 (最小) であるとき、以下のオイラー-ラグランジュ方程式が成り立つ

$$\frac{\partial G}{\partial y} = 0 \quad (59)$$

- 汎関数の例 (3)

- 今度は、 $y'(x)$ と x のみによって決まり、 $y(x)$ には依存しない関数 $G(y'(x), x)$ を考えよう
- この場合も変分 $\delta F / \delta y$ を求めるのは簡単である
- $G(y(x), y'(x), x)$ の変分の式に、 $\partial G / \partial y = 0$ を代入すればよい

$$\frac{\delta F}{\delta y(x)} = -\frac{d}{dx} \left(\frac{\partial G}{\partial y'} \right) \quad (60)$$

- オイラー-ラグランジュ方程式は次のようになる

$$-\frac{d}{dx} \left(\frac{\partial G}{\partial y'} \right) = 0 \quad (61)$$

● 汎関数の例 (4)

- $y(x)$ と $y'(x)$ によって決まる関数 $G(y(x), y'(x))$ を考えよう
- このときのオイラー-ラグランジュ方程式を導出してみよう
- $G(y(x), y'(x))$ を x で微分すれば

$$\begin{aligned} & \frac{d}{dx} G(y, y') \\ = & \frac{\partial}{\partial y} G(y, y') \frac{dy}{dx} + \frac{\partial}{\partial y'} G(y, y') \frac{dy'}{dx} \end{aligned} \quad (62)$$

$$= y' \frac{\partial}{\partial y} G(y, y') + \frac{\partial}{\partial y'} G(y, y') \frac{dy'}{dx} \quad (63)$$

となるから

$$y' \frac{\partial}{\partial y} G(y, y') = \frac{d}{dx} G(y, y') - \frac{\partial}{\partial y'} G(y, y') \frac{dy'}{dx} \quad (64)$$

また、オイラー-ラグランジュ方程式の両辺に y' を掛けたものは

$$y' \frac{\partial}{\partial y} G(y, y') - y' \frac{d}{dx} \left(\frac{\partial}{\partial y'} G(y, y') \right) = 0 \quad (65)$$

これらを連立させて

$$y' \frac{d}{dx} \left(\frac{\partial}{\partial y'} G(y, y') \right) = \frac{d}{dx} G(y, y') - \frac{\partial}{\partial y'} G(y, y') \frac{dy'}{dx} \quad (66)$$

$$y' \frac{d}{dx} \left(\frac{\partial}{\partial y'} G(y, y') \right) + \frac{\partial}{\partial y'} G(y, y') \frac{dy'}{dx} = \frac{d}{dx} G(y, y') \quad (67)$$

$$\frac{d}{dx} \left(y' \cdot \frac{\partial}{\partial y'} G(y, y') \right) = \frac{d}{dx} G(y, y') \quad (68)$$

$$\int \left(\frac{d}{dx} \left(y' \cdot \frac{\partial}{\partial y'} G(y, y') \right) \right) dx = \int \left(\frac{d}{dx} G(y, y') \right) dx + C \quad (69)$$

$$G(y, y') = y' \cdot \frac{\partial}{\partial y'} G(y, y') + C \quad (70)$$

となるので、結局オイラー-ラグランジュ方程式は

$$G - y' \frac{\partial G}{\partial y'} = \text{Const.} \quad (71)$$

と書ける

変分法で解ける問題の例

- 変分法で解ける問題の例 (1)

- 2 点 $P(0, 0)$ 、 $Q(a, b)$ を結ぶ最短経路は?
- 2 点を結ぶ経路 $y = f(x) (0 \leq x \leq a)$ の長さ l は、次のようになる

$$l = \int_0^a \sqrt{1 + \left(\frac{dy}{dx}\right)^2} dx = \int_0^a \sqrt{1 + y'^2} dx \quad (72)$$

- 被積分項が y' のみの関数となっていることが分かる

変分法で解ける問題の例

- $G(y'(x), x)$ の場合の公式を使えば、 l の $y = f(x)$ による変分が求まる

$$\frac{\delta l}{\delta f(x)} = -\frac{d}{dx} \left(\frac{\partial}{\partial y'} \sqrt{1 + y'^2} \right) \quad (73)$$

$$= -\frac{d}{dx} \left(\frac{1}{2} \frac{1}{\sqrt{1 + y'^2}} \frac{\partial}{\partial y'} (1 + y'^2) \right) \quad (74)$$

$$= -\frac{d}{dx} \frac{y'}{\sqrt{1 + y'^2}} \quad (75)$$

- l を最小化するような $y = f(x)$ は、上式の変分を 0 と等置すれば

$$-\frac{d}{dx} \frac{y'}{\sqrt{1 + y'^2}} = 0 \quad (76)$$

$$\therefore \frac{y'}{\sqrt{1 + y'^2}} = \text{Const.} \quad (77)$$

変分法で解ける問題の例

- これは、 y' が定数であることを意味している
- 従って、 $y = C_0x + C_1$ と書ける
- 以上より、2 点間を結ぶ最短経路は直線である
- $y = f(x)$ の形について、具体的な仮定は特に置いていないことに注意
- 変分法では、関数そのものを最適化する
- 従って、関数の具体的な形については、特に仮定する必要がない

変分法のまとめ

- 変分のまとめ

- これまでの計算で、次の変分が明らかとなった

$$\frac{\delta}{\delta y(x)} \int G(y(x), y'(x), x) dx = \frac{\partial}{\partial y} G(y(x), y'(x), x) - \frac{d}{dx} \left(\frac{\partial}{\partial y'} G(y(x), y'(x), x) \right) \quad (78)$$

$$\frac{\delta}{\delta y(x)} \int G(y(x), x) dx = \frac{\partial}{\partial y} G(y(x), x) \quad (79)$$

$$\frac{\delta}{\delta y(x)} \int G(y'(x), x) dx = -\frac{d}{dx} \left(\frac{\partial}{\partial y'} G(y'(x), x) \right) \quad (80)$$

変分法のまとめ

- ここまでの話の流れ

- 1 変分の定義や、変分の計算法について調査した
- 2 変分 (汎関数の微分) とは、入力関数が微小に変化したときの、出力値の変化量として定義される
- 3 汎関数が特定の形で表せるとき、変分がどのようなになるか計算した
- 4 汎関数が最大 (最小) になるとき、オイラー-ラグランジュ方程式が成立した
- 5 変分法を用いて、2 点間を結ぶ最短経路が直線になることを確認した

- これからの話の流れ

- 変分最適化を、どのように推論問題に適用するのかについて調べていく

- 変分推論が必要だった理由

- 潜在変数に関する事後分布 $p(\mathbf{Z}|\mathbf{X}, \theta)$ の計算は、困難であることが多い
- $p(\mathbf{Z}|\mathbf{X}, \theta)$ の厳密な計算は諦める代わりに、別の確率分布で近似したい
- 別の確率分布で近似するとき、単純な項の積として表現できるといった、何らかの仮定を置く

- 変分推論の目的

- 同時分布 $p(\mathbf{X}, \mathbf{Z})$ が求まっているときに、事後分布 $p(\mathbf{Z}|\mathbf{X})$ と、エビデンス $p(\mathbf{X})$ の近似を求める

- 注意点

- 観測変数と潜在変数をまとめて、 \mathbf{X}, \mathbf{Z} とおく
- $p(\mathbf{X})$ は、確率モデルからデータ \mathbf{X} が生起する確率である
- データからみたモデルの好みと解釈できるから、 $p(\mathbf{X})$ をモデルエビデンスという