

Abstract

Contrastive Divergence method における、標本数の極限をとった場合の収束性、推定精度 (分散の評価) について分かった結果を簡単にまとめました。途中の計算は丁寧に書いてありますが、内容としては少ないです。

まずは CD の収束性の証明です。(CD 法自体は Hinton によって 2002 年に発見されましたが、遅くとも 2005 年の論文 [1] では収束性について議論している論文がありました。しかしながら、この論文では収束先が真のパラメータに一致する事を示していますが、likelihood の curvature については論じてないです)。

0.1 推定精度の計算

以下では、経験分布の標本数極限について CD 法のパラメータの 1 階と 2 階の勾配を求めます。1 階については 0 に一致し、2 階については比熱のような項が生じる事を確認します。

計算の詳細に入る前に、略記した標本の集合と、確率分布を以下にまとめます。

$$p^{(0)}(x) := p^{(0)}(x|x^{(0)}) = \frac{1}{N} \sum_{n=1}^N \delta(x, x^{(0),n}) \quad (1)$$

$$p^{(1)}(x|J) := \sum_{x'} p(x|x'; J) p^{(0)}(x') \quad (2)$$

$$p^{(\infty)}(x|J) := \frac{e^{-Jg(x)}}{Z(J)} \quad (3)$$

$$p(x|x'; J) := \frac{e^{-Jg(x)}}{e^{-Jg(x)} + e^{-Jg(x')}} \quad (4)$$

$$g(x) := - \sum_{i=1}^d x_i x_{i+1} \quad (5)$$

$x^{(0)}$ を標本の一つとし、上の確率分布を順に、経験分布と、経験分布から一度遷移した分布とします。

$$\frac{\partial}{\partial J} CD(p^{(0)} \| p^{(1)}(|J)) = \frac{\partial}{\partial J} \left(\sum_{x:all} p^{(0)}(x) \log \frac{p^{(0)}(x)}{p^{(\infty)}(x|J)} - \sum_{x:all} p^{(1)}(x|J) \log \frac{p^{(1)}(x|J)}{p^{(\infty)}(x|J)} \right) \quad (6)$$

$$= - \sum_{x:all} (p^{(0)}(x) - p^{(1)}(x|J)) \frac{\partial}{\partial J} \log p^{(\infty)}(x|J) - \sum_{x:all} p^{(1)}(x|J) \frac{\partial}{\partial J} \log p^{(1)}(x|J) \quad (7)$$

$$= - \sum_{x:all} (p^{(0)}(x) - p^{(1)}(x|J)) \frac{\partial}{\partial J} \log p^{(\infty)}(x|J) - \frac{\partial}{\partial J} \sum_{x:all} p^{(1)}(x|J) \quad (8)$$

$$(9)$$

ここで、第一項目の $(p^{(0)}(x) - p^{(1)}(x|J))$ は経験分布について平均をとり、パラメータ J を真のパラメータにセットすると消えます。すなわち

$$[\langle p^{(1)}(x|J) \rangle_{x^{(0)}|J_0}]_{J=J_0} = [\langle p^{(1)}(x|J) \rangle_{x^{(0)}|J_0}]_{J=J_0} \quad (10)$$

$$= \left[\sum_{x^{(0)}} p(x|x^{(0)}; J) p^{(\infty)}(x^{(0)}|J_0) \right]_{J=J_0} \quad (11)$$

$$= p^{(\infty)}(x|J_0) \quad (12)$$

かつ

$$[\langle p^{(0)}(x) \rangle_{x^{(0)}|J_0}]_{J=J_0} = [\langle p^{(0)}(x|x^{(0)}) \rangle_{x^{(0)}|J_0}]_{J=J_0} \quad (13)$$

$$= p^{(\infty)}(x|J_0) \quad (14)$$

また、第二項目は遷移後も確率分布になっているためパラメータ微分で消えます。
ゆえに、

$$[\langle \frac{\partial}{\partial J} CD(p^{(0)} \| p^{(1)}(|J)) \rangle_{x^{(0)}|J_0}]_{J=J_0} = 0 \quad (15)$$

ただし、

$$\langle \cdot \rangle_{x^{(0)}|J_0} := \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \cdot \delta(x, x^{(0),n}) = \sum_x \cdot p^{(\infty)}(x|J_0) \quad (16)$$

ここで、 $x^{(0),n}$ は n 番目の標本で、真のパラメータ J_0 に従ってサンプリングされているとしています。

likelihood の平均の極値は、真のパラメータに一致する事がわかりました。

$p^{(infy)}(x|\theta)$ が J について指数分布的 $e.i.p^{(\infty)}(x|\theta) \propto e^{-\theta \cdot u(x)}$ であるならば、CD の評価関数は parameter J について凸関数になることが知られています。凸性と、上の議論における極値が真のパラメータに一致する結果により、任意の初期状態から最適化を開始しても大域的な最適解に収束することがわかりました!!

CD 法の手足性については 2005 年以降から幾つかの論文で議論されており、同様に真のパラメータへの収束を示していました (Alan Yuille, 2005)

次に CD の評価関数の 2 階微分について経験分布の平均を取り、パラメータを真のパラメータにセットした量が比熱のような量としてかけることを示します。

$$\frac{\partial^2}{\partial J^2} CD = \frac{\partial}{\partial J} \left\{ \sum_{x:all} (p^{(1)}(x|J) - p^{(0)}(x)) \frac{\partial}{\partial J} \log p^{(\infty)}(x|J) - \frac{\partial}{\partial J} \sum_{x:all} p^{(1)}(x|J) \right\} \quad (17)$$

$$\begin{aligned} &= \sum_x \frac{\partial p^{(1)}(x|J)}{\partial J} \frac{\partial}{\partial J} \log p^{(\infty)}(x|J) + \sum_x (p^{(1)}(x|J) - p^{(0)}(x)) \frac{\partial^2}{\partial J^2} \log p^{(\infty)}(x|J) \\ &- \sum_x \frac{\partial^2 p^{(1)}(x|J)}{\partial J^2} \log \frac{p^{(1)}(x|J)}{p^{(\infty)}(x|J)} - \sum_x \frac{\partial p^{(1)}(x|J)}{\partial J} \left(\frac{\partial}{\partial J} \log p^{(1)}(x|J) - \frac{\partial}{\partial J} p^{(\infty)}(x|J) \right) \end{aligned} \quad (18)$$

ここで、一階微分の場合と同様に、経験分布について平均を取り、パラメータを真のパラメータにセットすると第一項目を除いて消えてしまいます。従って二階の微分は結局

$$[\langle \frac{\partial^2}{\partial J^2} CD(p^{(0)} \| p^{(1)}(|J)) \rangle_{x^{(0)}|J_0}]_{J=J_0} = [\langle \sum_x \frac{\partial p^{(1)}(x|J)}{\partial J} \frac{\partial}{\partial J} \log p^{(\infty)}(x|J) \rangle_{x^{(0)}|J_0}]_{J=J_0} \quad (19)$$

上式の $p^{(1)}(x|J)$ のパラメータ微分の平均の扱いが厄介ではありますが、詳細に踏み込まず、もう少し簡単な形にできます。

平衡分布は以下のように形で与えられていることから

$$\frac{\partial}{\partial J} \log p^{(1)}(x|J) = \frac{\partial}{\partial J} \log \frac{e^{-Jg(x)}}{Z(J)} \quad (20)$$

$$= -g(x) + \langle g(x) \rangle_{x|J} \quad (21)$$

ここで、 $\langle g(x) \rangle_{x|J}$ は x に依らないため、CD の二階微分の初期分布についての平均に上式を入れた場合、以下のように 0 となります。

$$[\langle \sum_x \frac{\partial p^{(1)}(x|J)}{\partial J} \langle g(x) \rangle_{x|J} \rangle_{x^{(0)}|J_0}]_{J=J_0} = [\langle g(x) \rangle_{x|J} \langle \sum_x \frac{\partial p^{(1)}(x|J)}{\partial J} \rangle_{x^{(0)}|J_0}]_{J=J_0} \quad (22)$$

$$= [\langle g(x) \rangle_{x|J} \langle 0 \rangle_{x^{(0)}|J_0}]_{J=J_0} = 0 \quad (23)$$

さらに計算を実行していくと

$$[\langle \frac{\partial^2}{\partial J^2} CD(p^{(0)} \| p^{(1)}(|J)) \rangle_{x^{(0)}|J_0}]_{J=J_0} \quad (24)$$

$$= -[\langle \sum_x \frac{\partial p^{(1)}(x|J)}{\partial J} g(x) \rangle_{x^{(0)}|J_0}]_{J=J_0} \quad (25)$$

$$= -[\sum_x g(x) \langle \frac{\partial p^{(1)}(x|J)}{\partial J} \rangle_{x^{(0)}|J_0}]_{J=J_0} \quad (26)$$

$$= -[\sum_x g(x) \sum_{x^{(0)}} \frac{\partial p(x|x^{(0)}; J)}{\partial J} p^{(\infty)}(x^{(0)}|J_0)]_{J=J_0} \quad (27)$$

$$= -\sum_x g(x) \sum_{x^{(0)}} \frac{\partial p(x|x^{(0)}; J)}{\partial J} |_{J=J_0} p^{(\infty)}(x^{(0)}|J_0) \quad (28)$$

$$= -\sum_x g(x) \sum_{x^{(0)}} \frac{\partial p(x|x^{(0)}; J)}{\partial J} |_{J=J_0} (p^{(\infty)}(x^{(0)}|J) + (p^{(\infty)}(x^{(0)}|J_0) - p^{(\infty)}(x^{(0)}|J))) \quad (29)$$

$$= -\sum_x g(x) \frac{\partial}{\partial J} |_{J=J_0} \sum_{x^{(0)}} p(x|x^{(0)}; J) \{p^{(\infty)}(x^{(0)}|J) + (p^{(\infty)}(x^{(0)}|J_0) - p^{(\infty)}(x^{(0)}|J))\} \quad (30)$$

$$= -\sum_x g(x) \frac{\partial}{\partial J} |_{J=J_0} \{p^{(\infty)}(x|J) + \sum_{x^{(0)}} p(x|x^{(0)}; J) (p^{(\infty)}(x^{(0)}|J_0) - p^{(\infty)}(x^{(0)}|J))\} \quad (31)$$

$$= -\sum_x g(x) \{-g(x)p^{(\infty)}(x|J_0) + \langle g(x) \rangle_{x|J_0} p^{(\infty)}(x|J_0)\} \quad (32)$$

$$- \sum_x g(x) \frac{\partial}{\partial J} |_{J=J_0} \sum_{x^{(0)}} p(x|x^{(0)}; J) \{p^{(\infty)}(x^{(0)}|J_0) - p^{(\infty)}(x^{(0)}|J)\} \quad (33)$$

$$= -(\langle g(x)^2 \rangle_{x|J_0} - \langle g(x) \rangle_{x|J_0}^2) \quad (34)$$

$$- \sum_x g(x) \frac{\partial}{\partial J} |_{J=J_0} \sum_{x^{(0)}} p(x|x^{(0)}; J) \{p^{(\infty)}(x^{(0)}|J_0) - p^{(\infty)}(x^{(0)}|J)\} \quad (35)$$

$$= -(\langle g(x)^2 \rangle_{x|J_0} - \langle g(x) \rangle_{x|J_0}^2) - \sum_x g(x) \frac{\partial}{\partial J} |_{J=J_0} \sum_{x^{(0)}} p(x|x^{(0)}; J) p^{(\infty)}(x^{(0)}|J_0) \quad (36)$$

第一項目は明らかに比熱に比例するような項です。第二項目は補正項の役割を果たしています (比熱を減少させるか、増加させるかを知る必要があります)。

0.2 Correction の解析

補正の計算が実行できました。

$$Correction = \sum_x g(x) \frac{\partial}{\partial J} |_{J=J_0} \sum_{x^{(0)}} p(x|x^{(0)}; J) [p^{(\infty)}(x^{(0)}|J) - p^{(\infty)}(x^{(0)}|J_0)] \quad (37)$$

$$= \sum_x g(x) \sum_{x^{(0)}} \frac{\partial p(x|x^{(0)}; J)}{\partial J} \Big|_{J=J_0} [p^{(\infty)}(x^{(0)}|J) - p^{(\infty)}(x^{(0)}|J_0)] \quad (38)$$

$$+ \sum_x g(x) \sum_{x^{(0)}} p(x|x^{(0)}; J) \frac{\partial}{\partial J} \Big|_{J=J_0} [p^{(\infty)}(x^{(0)}|J) - p^{(\infty)}(x^{(0)}|J_0)]$$

$$= 0 + \sum_x g(x) \sum_{x^{(0)}} p(x|x^{(0)}; J) \frac{\partial}{\partial J} \Big|_{J=J_0} [p^{(\infty)}(x^{(0)}|J) - 0] \quad (39)$$

$$= \sum_x g(x) \sum_{x^{(0)}} p(x|x^{(0)}; J) \{-g(x) + \langle g(x) \rangle_{x|J_0}\} p^{(\infty)}(x^{(0)}|J_0) \quad (40)$$

$$= \sum_x g(x) \{-g(x) + \langle g(x) \rangle_{x|J}\} \sum_{x^{(0)}} p(x|x^{(0)}; J) p^{(\infty)}(x^{(0)}|J_0) \quad (41)$$

$$= \sum_x g(x) \{-g(x) + \langle g(x) \rangle_{x|J}\} p^{(\infty)}(x|J_0) \quad (42)$$

$$= -(\langle g(x)^2 \rangle_{x|J_0} - \langle g(x) \rangle_{x|J_0}^2) \quad (43)$$

なんだか、妙な気もしますが、前章の結果と合わせると

$$[\langle \frac{\partial^2}{\partial J^2} CD(p^{(0)} \| p^{(1)}(|J)) \rangle_{x^{(0)}|J_0}]_{J=J_0} = -2(\langle g(x)^2 \rangle_{x|J_0} - \langle g(x) \rangle_{x|J_0}^2) \quad (44)$$

上式の結果を利用すると Hessian は殆ど解けた事になります。

$$H[CD] := [\langle \frac{\partial}{\partial \vec{J}} \frac{\partial}{\partial \vec{J}} CD(p^{(0)} \| p^{(1)}(|J)) \rangle_{x^{(0)}|J_0}]_{J=J_0} \quad (45)$$

$$= -2(\langle \vec{g}(x) \vec{g}(x)^T \rangle_{x|J_0} - \langle \vec{g}(x) \rangle_{x|J_0} \langle \vec{g}(x) \rangle_{x|J_0}^T) \quad (46)$$

ただし、 $\frac{\partial}{\partial \vec{J}}$ を縦ベクトルで定義しています。

ここで、任意のベクトル \vec{a} を用いて Hessian の正定値性条件が確認できます。

$$\vec{a}^T H[CD] \vec{a} = \sum_i^d \sum_j^d \sum_x p^{(\infty)}(x|J_0) (x) a_i a_j g_i(x) (-g_j(x) + \langle g_j(x) \rangle_{x|J_0}) \quad (47)$$

$$= \sum_i^d \sum_j^d a_i a_j \langle g_i(x) (-g_j(x) + \langle g_j(x) \rangle_{x|J_0}) \rangle_{x|J_0} \quad (48)$$

$g(x)$ の具体的な標識を与えることで共分散行列の i, j 要素を評価します。ここで、 $g(x) = -\sum_{i=1}^d x_i x_{i+1}$ で与えられている場合は

$$\langle g_i(x) \rangle_{x|J_0} \langle g_j(x) \rangle_{x|J_0} - \langle g_i(x) g_j(x) \rangle_{x|J_0} = \begin{cases} 0 & \text{for } |i - j| > 1 \\ C_{ij}(J_0) & \text{for } |i - j| \leq 1 \end{cases} \quad (49)$$

となり、Ising Model における磁場無しの $\langle g(x) \rangle_{x|J_0}$ は 0 であり、 $\langle g(x)^2 \rangle_{x|J_0} < 0$ より、分散は正になっています。

従って (半) 正定値性が示され、凸性がわかります (あっ、けどこれ真のパラメータ周りでの自明な凸性しか言っていないですよ。。。)。

0.3 MPF の場合

CD 法で調べた量を MPF に対しても同様に議論します。

$$\langle \frac{\partial}{\partial J} MPF \rangle_{x^{(0)}|J_0} = \langle \frac{\partial}{\partial J} |_{J=J_0} \sum_x p^{(0)}(x) \log \frac{p^{(0)}(x)}{p^{(\epsilon)}(x|J)} \rangle_{x^{(0)}|J_0} \quad (50)$$

$$= \langle \sum_x \frac{p^{(0)}(x)}{p^{(\epsilon)}(x|J)} \frac{\partial p^{(\epsilon)}(x|J)}{\partial J} |_{J=J_0} \rangle_{x^{(0)}|J_0} \quad (51)$$

$$= \sum_x \frac{p^{(0)}(x)}{p^{(\infty)}(x|J)} |_{J=J_0} = 0 \quad (52)$$

真のパラメータに収束することが分かります。また、分散についても同様に。

$$\langle \frac{\partial^2}{\partial J^2} MPF \rangle_{x^{(0)}|J_0} \quad (53)$$

$$= - \sum_x p^{(0)}(x) \{ \langle \frac{1}{p^{(\epsilon)}(x|x^{(0)}; J)} \frac{\partial^2 p^{(\epsilon)}(x|x^{(0)}; J)}{\partial J^2} \rangle_{x^{(0)}|J_0} - \langle \frac{1}{p^{(\epsilon)}(x|x^{(0)}; J)} \frac{\partial p^{(\epsilon)}(x|x^{(0)}; J)}{\partial J} \rangle_{x^{(0)}|J_0}^2 \} \quad (54)$$

$$= \langle g(x)^2 \rangle_{x|J_0} - \langle g(x) \rangle_{x|J_0}^2 \quad (55)$$

示されました。