林祐輔（所属: Japan Digital Design, 連絡先: bheprr@gmail.com, Twitter: @hayashiyus）

イントロ（1/n）

今日は，表現学習と熱力学の形式的類似性を紹介した後，その拘束条件について議論します．

そのお話をする前に，生成モデルの基礎についてお話します．

生成モデルと認識

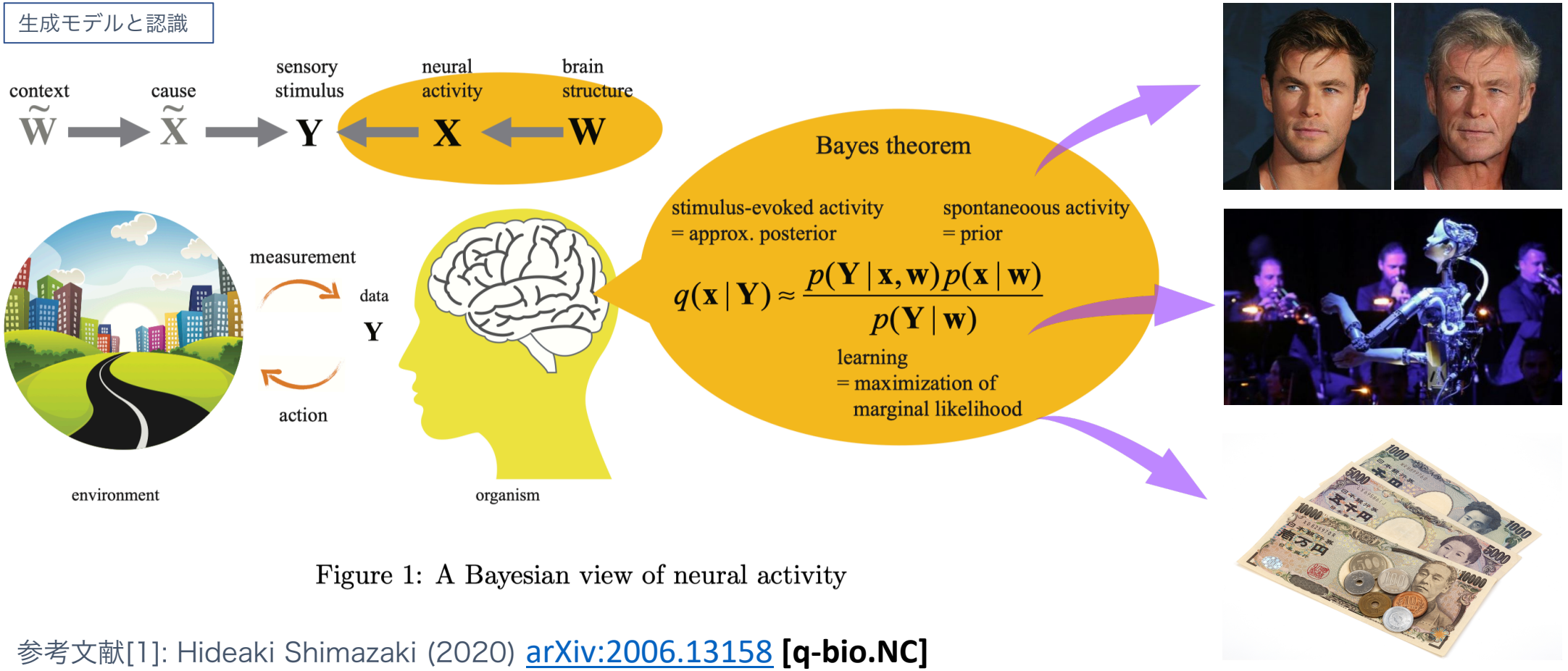


Figure 1: A Bayesian view of neural activity

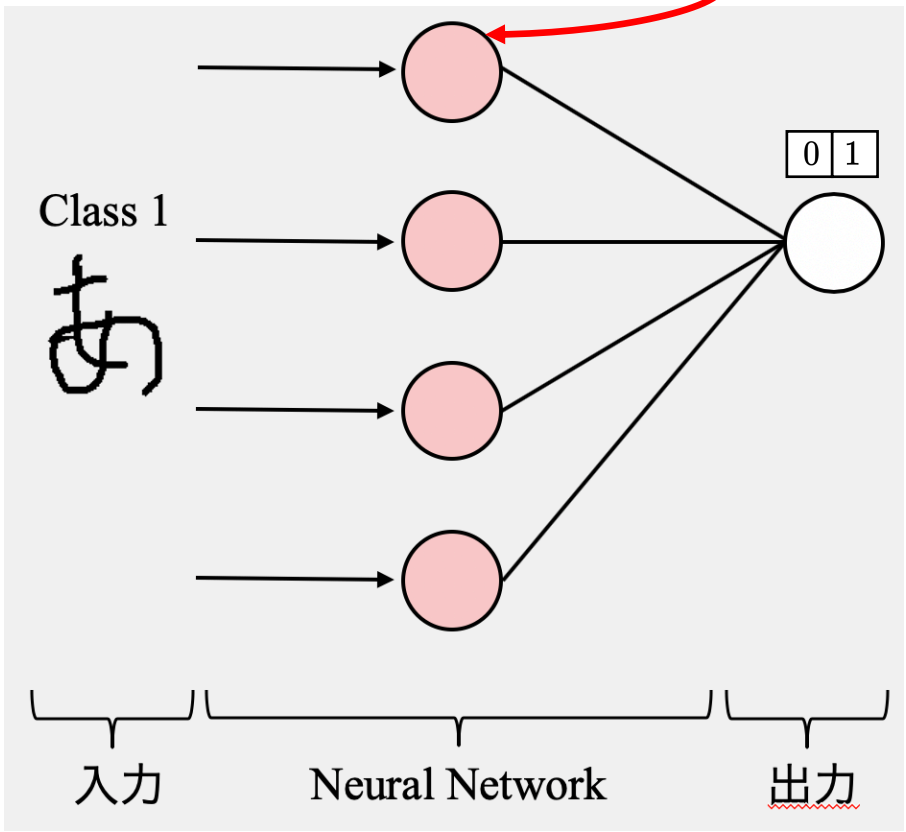
参考文献[1]: Hideaki Shimazaki (2020) arXiv:2006.13158 **[q-bio.NC]**

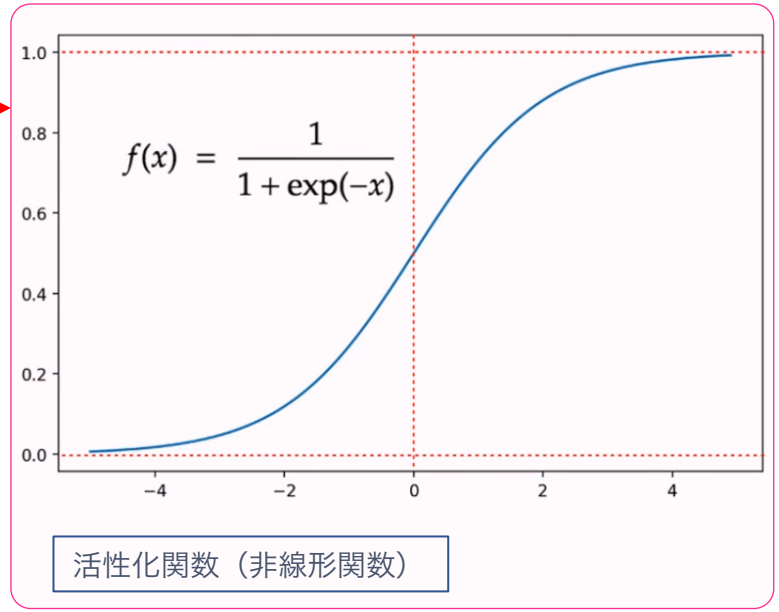
参考文献[2]: Sumio Watanabe, Mathematical Theory of Bayesian Statistics, CRC Press (2018)

林祐輔（所属: Japan Digital Design, 連絡先: bheprr@gmail.com, Twitter: @hayashiyus）

イントロ（2/n）

- 機械学習でやっていること
  - ➢ 入力 x と出力 y を結びつける関数 $f(x)$ を推定する.





$$\theta_{k+1} = \theta_k - \eta \frac{\partial \mathcal{L}(y'_k, y_k)}{\partial \theta_k}$$

$\mathcal{L}(y', y)$ 目的関数（誤差関数）

y' 0 1 真の教師ラベル

0 1 $y = f(\text{あ})$ モデルの予測

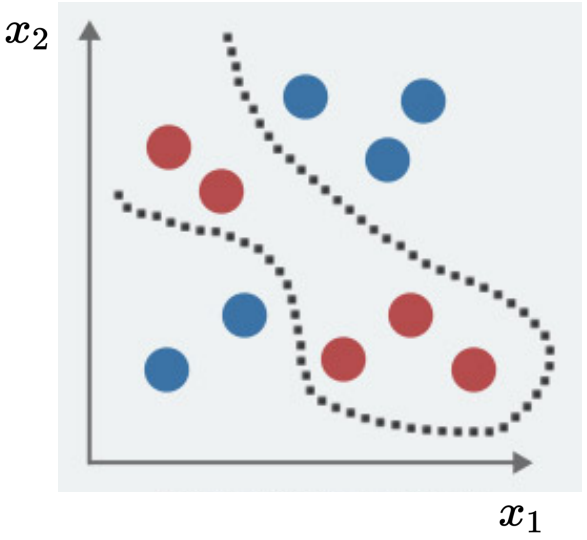
林祐輔（所属: Japan Digital Design, 連絡先: bheprr@gmail.com, Twitter: @hayashiyus）

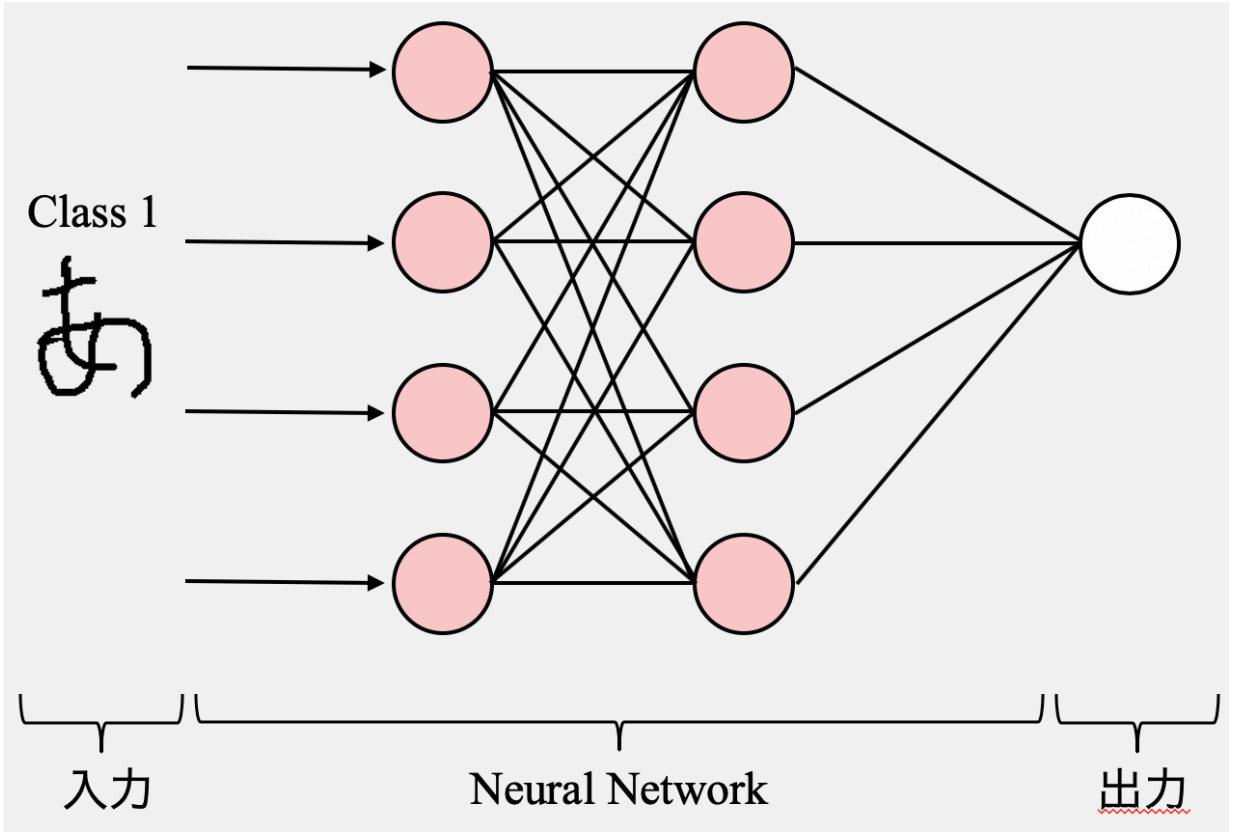
イントロ（3/n）

- ニューラルネットの層を一つ増やしたときに何が起こるか.

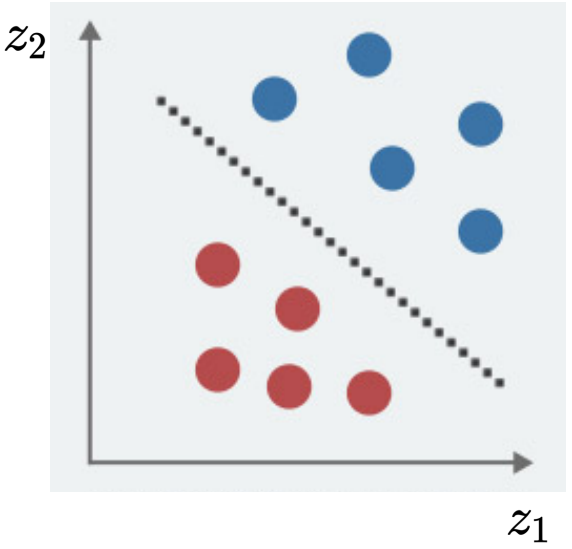
$$z = f_1(x) \quad y = f_2(z) = f_2(f_1(x))$$

線形分離不可能な複雑な分布





線形分離可能な単純な分布



- 統計モデルの内部では，入力 x を，より単純な分布をもつ特徴量 z に変換している.
  中間層無しのニューラルネットは，ロジスティック回帰モデルと同値.

林祐輔（所属: Japan Digital Design, 連絡先: bheprr@gmail.com, Twitter: @hayashiyus）

イントロ（4/n）

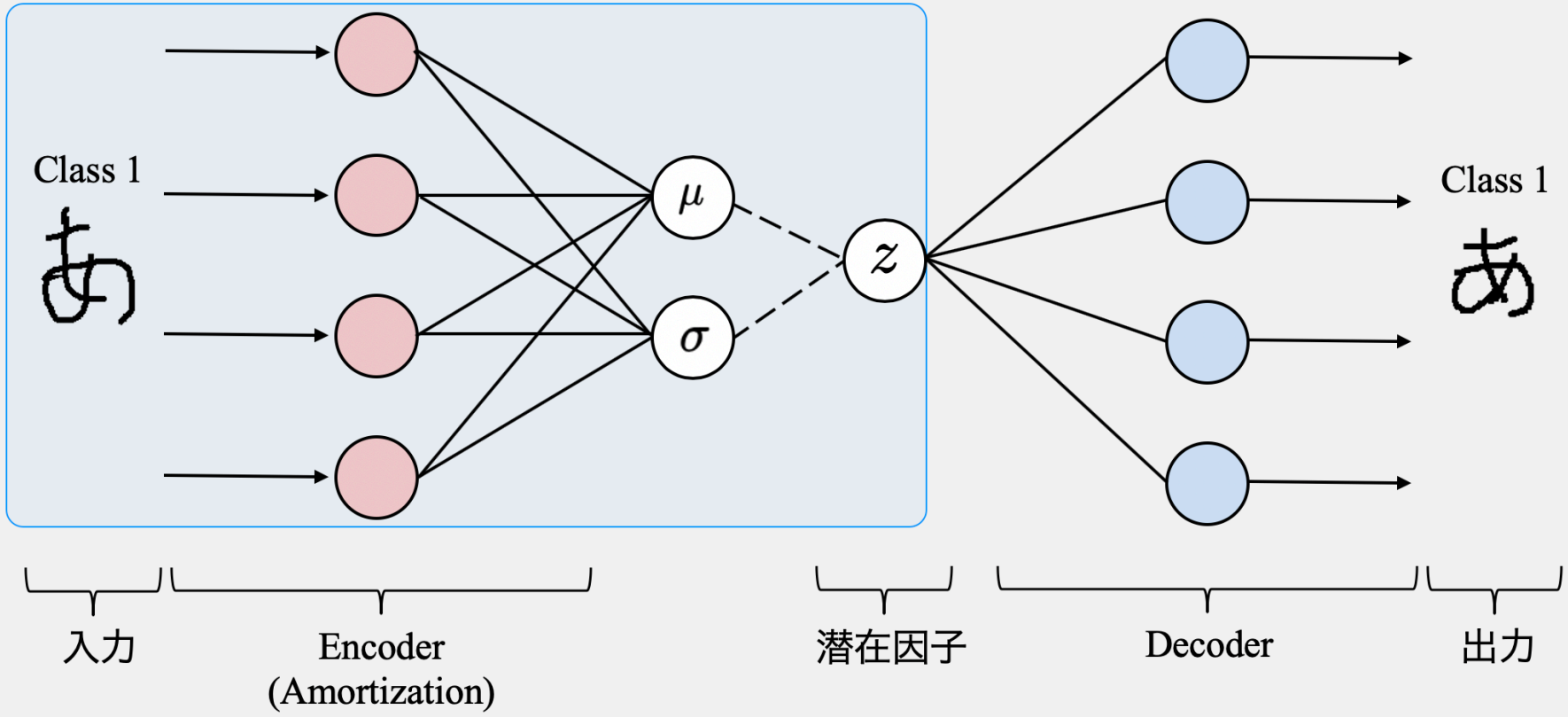
- 生成モデル（VAE）

$$z \sim \mathcal{N}\left(\mu(x), \sigma^2(x)\right)$$

ガウス分布

ニューラルネットがつくる確率関数

$$q(x, z) = q(x)q(z \mid x)$$



林祐輔（所属: Japan Digital Design, 連絡先: bheprr@gmail.com, Twitter: @hayashiyus）

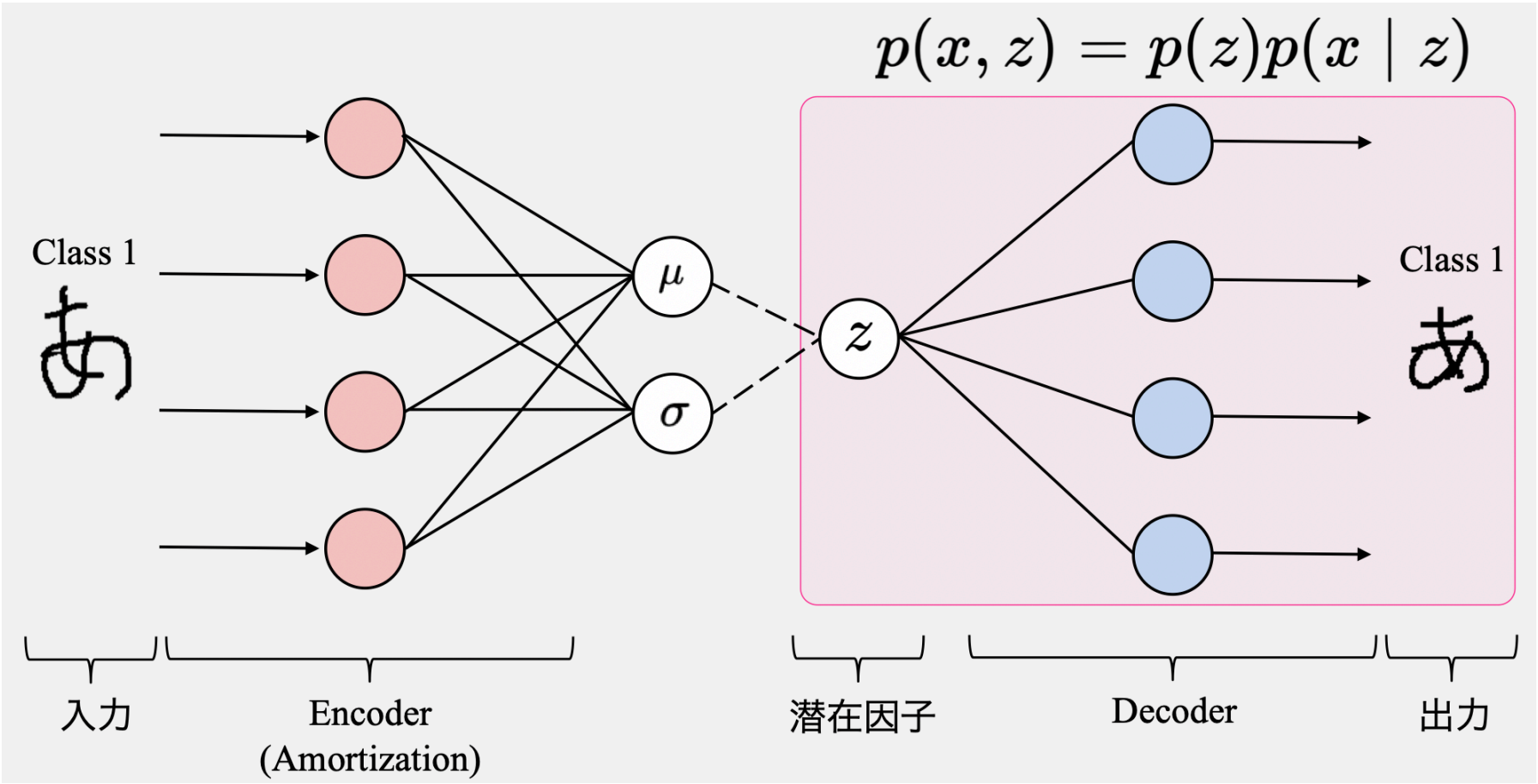
イントロ（5/n）

- 生成モデル（VAE）

ガウス分布 $z \sim \mathcal{N}(0, 1)$

ニューラルネットがつくる確率関数

$$p(x, z) = p(z)p(x \mid z)$$

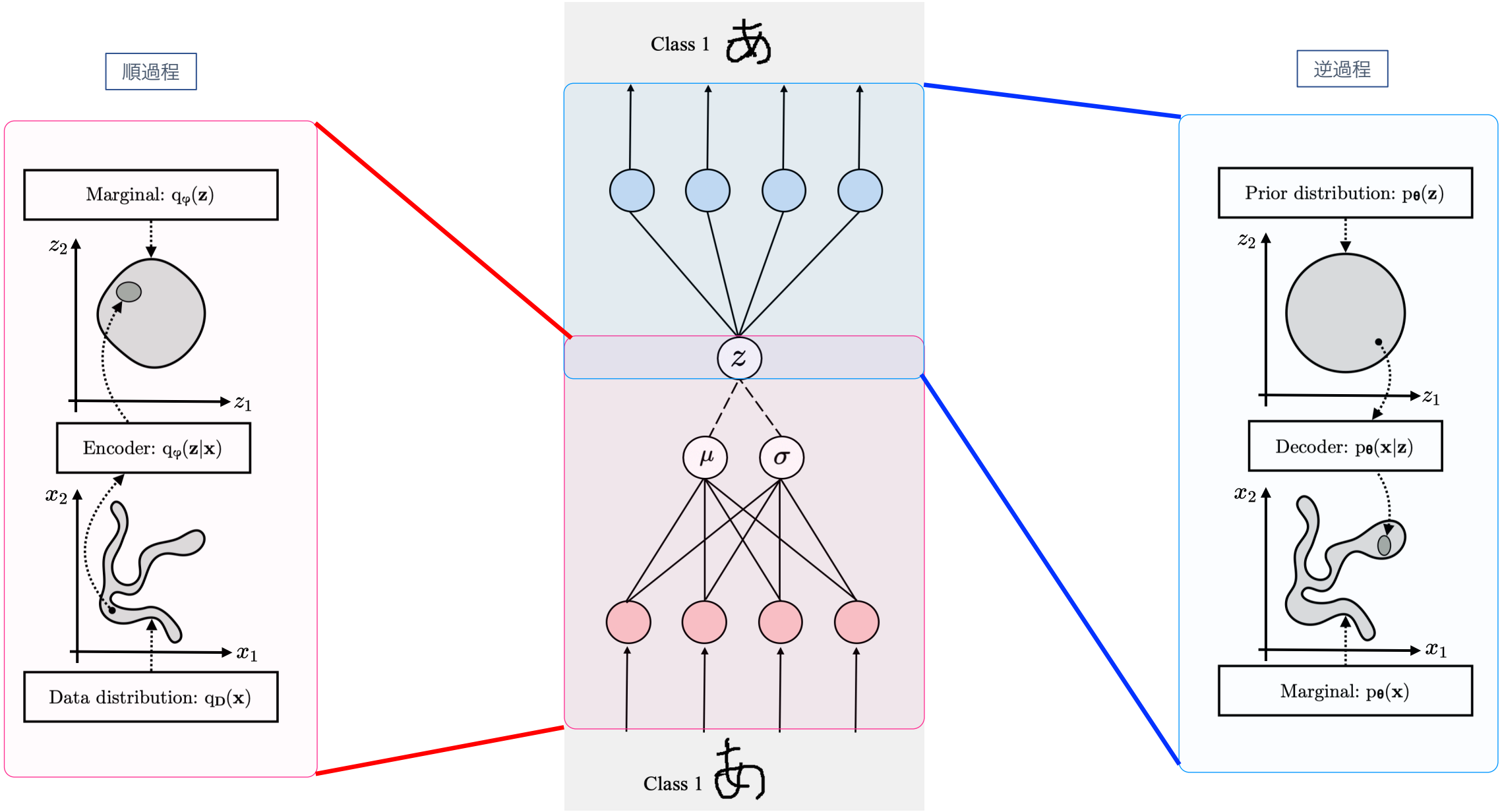


林祐輔（所属: Japan Digital Design, 連絡先: bheprr@gmail.com, Twitter: @hayashiyus）

イントロ（6/n）

順過程

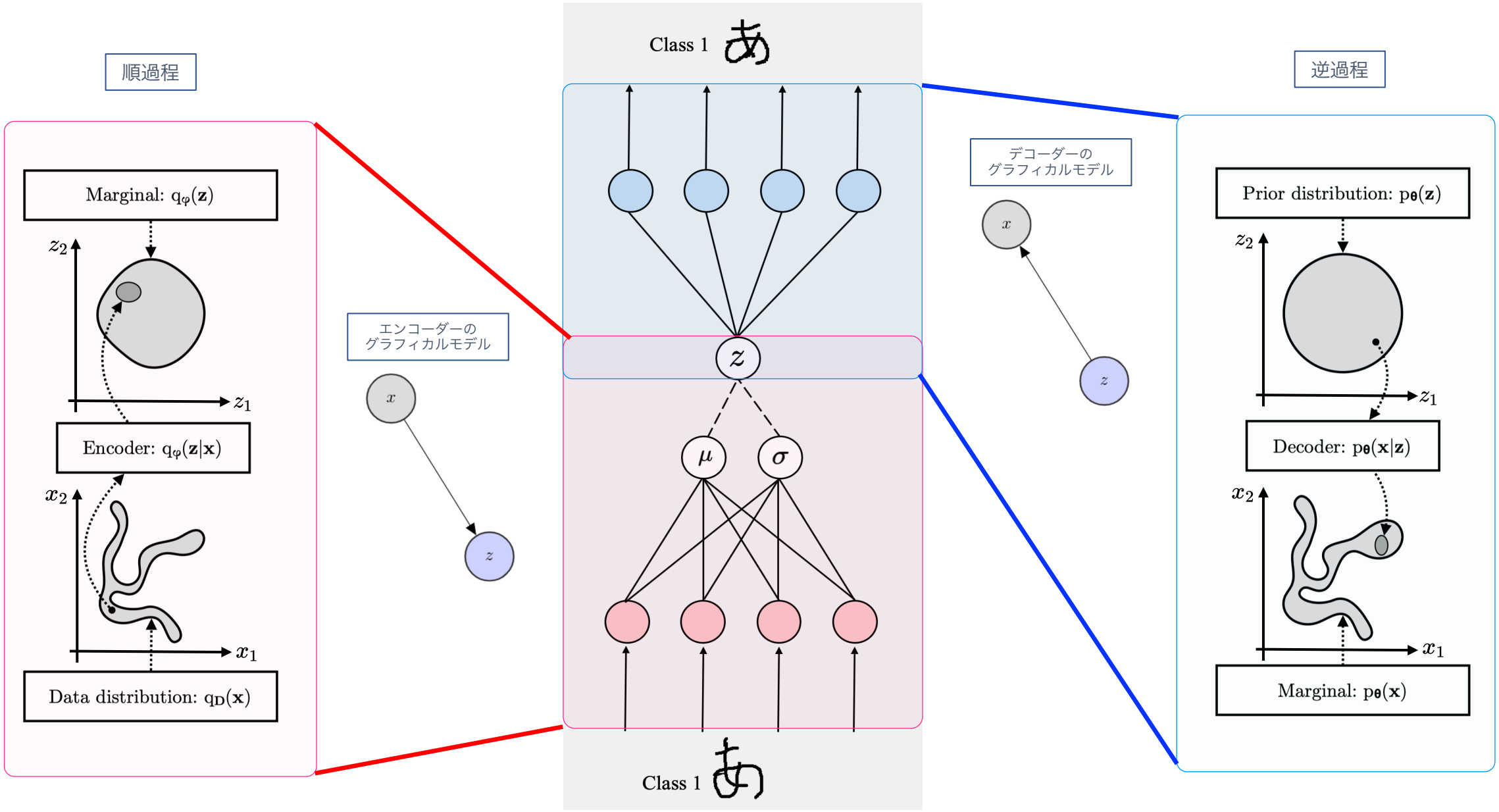
逆過程



参考文献[3]: Kingma and Welling (2019) arXiv:1906.02691 **[cs.LG]**

林祐輔（所属: Japan Digital Design, 連絡先: bheprr@gmail.com, Twitter: @hayashiyus）

イントロ（7/n）

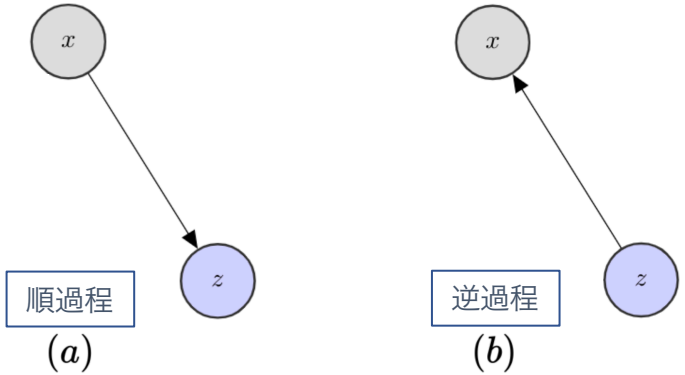


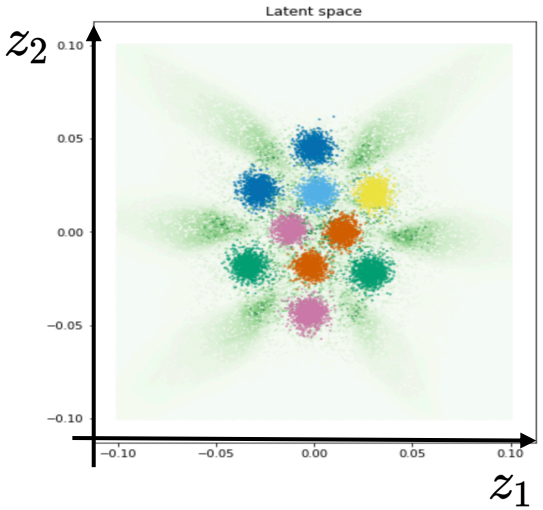
参考文献[3]: Kingma and Welling (2019) arXiv:1906.02691 **[cs.LG]**

林祐輔（所属: Japan Digital Design, 連絡先: bheprr@gmail.com, Twitter: @hayashiyus）

イントロ（8/n）

- 生成モデルの学習は，順過程と逆過程，それぞれの関数を推定するプロセスと理解できる．
  その目的関数（損失関数）は，順過程と逆過程，それぞれの同時分布を近づける形になっている．





$p(\boldsymbol{s}) := T(\boldsymbol{z},\boldsymbol{x})\,p(\boldsymbol{x}) = p(\boldsymbol{z}|\boldsymbol{x})\,p(\boldsymbol{x})$

$q(\boldsymbol{s}) := T(\boldsymbol{x},\boldsymbol{z})\,q(\boldsymbol{z}) = q(\boldsymbol{x}|\boldsymbol{z})\,q(\boldsymbol{z})$

$$\begin{aligned}\mathcal{D}(p(\boldsymbol{s})\|q(\boldsymbol{s})) &= \int p(\boldsymbol{s})\log\frac{p(\boldsymbol{s})}{q(\boldsymbol{s})}d\boldsymbol{s}\\ &= \int p(\boldsymbol{s})\log p(\boldsymbol{x})d\boldsymbol{s} + \int p(\boldsymbol{s})\log\frac{p(\boldsymbol{z}|\boldsymbol{x})}{q(\boldsymbol{z})}d\boldsymbol{s} - \int p(\boldsymbol{s})\log q(\boldsymbol{x}|\boldsymbol{z})d\boldsymbol{s}\\ &= -H(\boldsymbol{x}) + \underbrace{R + D}_{\mathcal{L}_{\mathrm{ELBO}}} \geq 0\end{aligned}$$

データのエントロピー

$$H(\boldsymbol{x}) := -\int p(\boldsymbol{x})\log p(\boldsymbol{x})d\boldsymbol{x}$$

VAEの目的関数 変分下限（ELBO）

$$\begin{aligned}\mathcal{D}(p(\boldsymbol{s})\|q(\boldsymbol{s})) &= \int p(\boldsymbol{s})\log\frac{p(\boldsymbol{s})}{q(\boldsymbol{s})}d\boldsymbol{s}\\ &= \int p(\boldsymbol{s})\log\frac{p(\boldsymbol{z}|\boldsymbol{x})p(\boldsymbol{x})}{q(\boldsymbol{x}|\boldsymbol{z})p(\boldsymbol{z})}d\boldsymbol{s} + \int p(\boldsymbol{s})\log\frac{p(\boldsymbol{z})}{q(\boldsymbol{z})}d\boldsymbol{s}\\ &:= \sigma^{\mathcal{S}}_{\mathrm{bath}} + \sigma^{\mathcal{S}}_{\mathrm{sys}} = \sigma^{\mathcal{S}}_{\mathrm{tot}} \geq 0\end{aligned}$$

全系のエントロピー生成率

この変分下限を**＜自由エネルギー＞**とみなす

エントロピー生成と自由エネルギーには

$\sigma^{\mathcal{S}}_{\mathrm{tot}} = \beta[W - \Delta F]$

の関係式が成り立つので

$\beta W = -H(\boldsymbol{x})$

$-\beta\Delta F = R + D$

を同一視すると約束すれば

$\sigma^{\mathcal{S}}_{\mathrm{tot}} = -H(\boldsymbol{x}) + R + D$

林祐輔（所属: Japan Digital Design, 連絡先: bheprr@gmail.com, Twitter: @hayashiyus）

表現学習の熱力学（9/n）

ここから表現学習と熱力学の形式的類似性について議論していきます．後半で，転移学習のプロセスに準静的過程のアナロジーを導入して，サイクルとその熱効率について議論します．

- 確率熱力学：ゆらぐ系の熱力学．微小系が状態 X から状態 Z に遷移する過程（順過程と呼ぶ）の確率を p とし，その逆過程の確率を q とするとき，p と q との距離が，微小系の状態変化に伴うエントロピー生成 σ_{tot} を決定する．

$$\mathcal{D}\left(p(\boldsymbol{s})||q(\boldsymbol{s})\right) = \mathcal{D}\left(p(\boldsymbol{s})||q^*(\boldsymbol{s})\right) + \mathcal{D}\left(q^*(\boldsymbol{s})||q(\boldsymbol{s})\right).$$

特徴量の生成プロセスの学習

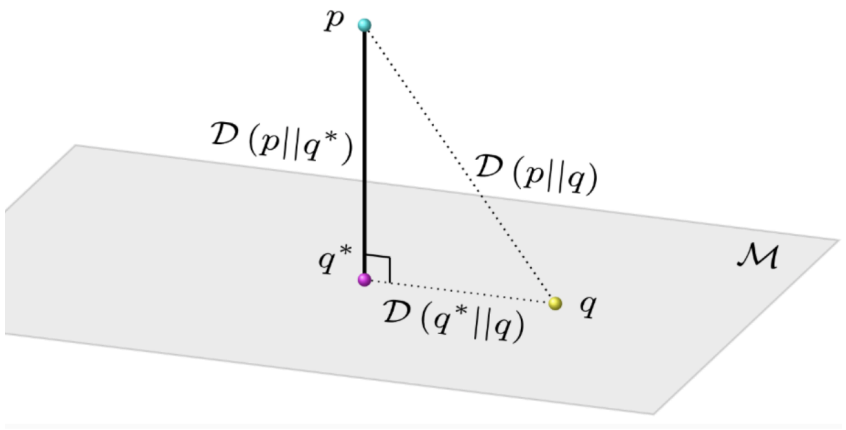


Figure 1: Information-Geometric Pythagorean theorem.

情報幾何として共通の幾何構造

$$\sigma^{\mathcal{S}}_{\mathrm{tot}} := \sigma^{\mathcal{S}}_{\mathrm{sys}} + \sigma^{\mathcal{S}}_{\mathrm{bath}}.$$

ゆらぐ系のエントロピー生成

entropy production (2nd law)

$\sigma^{\mathcal{S}}_{\mathrm{tot}} \ (\geq 0)$

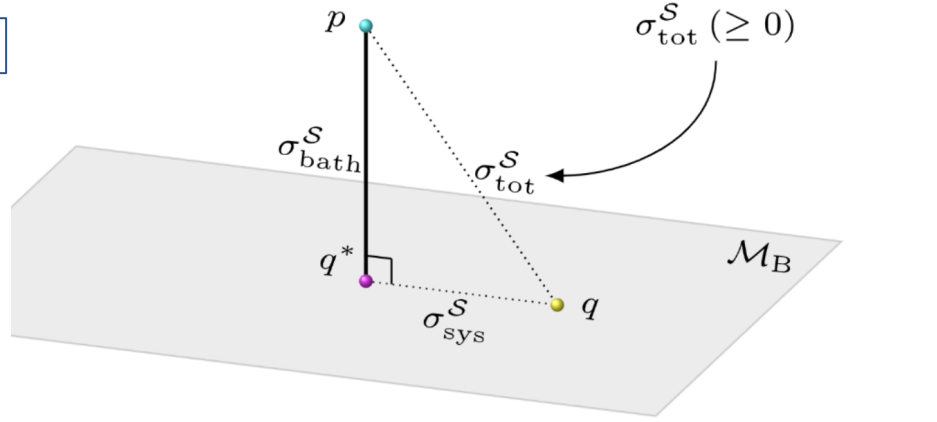


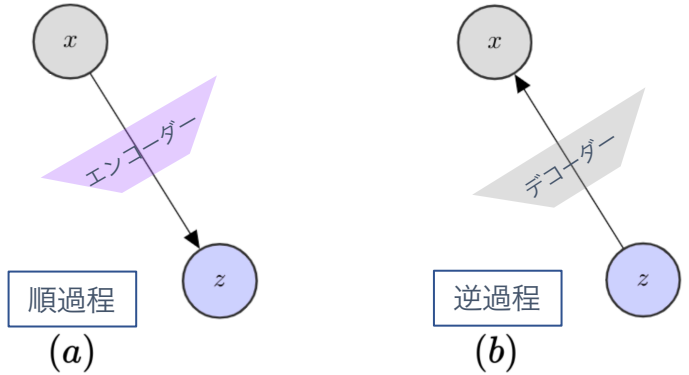
Figure 2: Total entropy production.

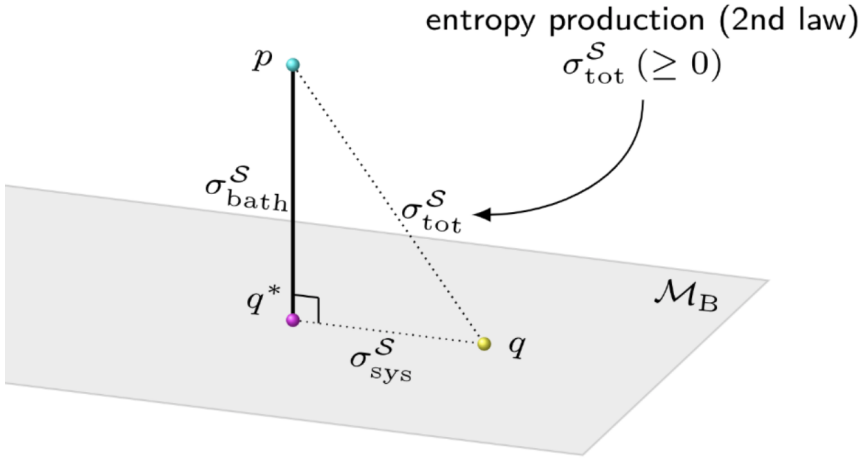
参考文献[4]: Ito, Oizumi and Amari (2020) arXiv:1810.09545v5 **[cond-mat.stat-mech]**

林祐輔（所属: Japan Digital Design, 連絡先: bheprr@gmail.com, Twitter: @hayashiyus）

表現学習の熱力学（10/n）

- 生成モデルの単純なモデル：変分オートエンコーダー（VAE）を考える.





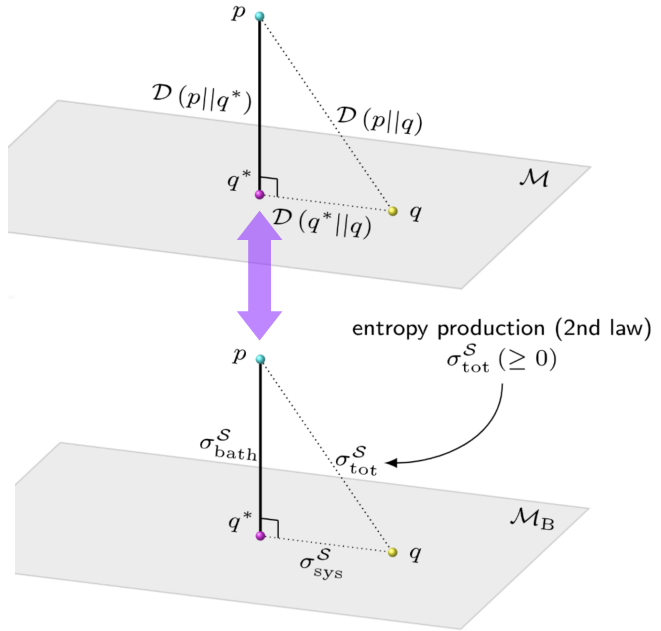
$p(\boldsymbol{s}) := T(\boldsymbol{z},\boldsymbol{x})\,p(\boldsymbol{x}) = p(\boldsymbol{z}|\boldsymbol{x})\,p(\boldsymbol{x})$.

$q(\boldsymbol{s}) := T(\boldsymbol{x},\boldsymbol{z})\,q(\boldsymbol{z}) = q(\boldsymbol{x}|\boldsymbol{z})\,q(\boldsymbol{z})$

$\mathcal{M}_{\rm B} = \{q(\boldsymbol{s}) \mid q(\boldsymbol{s}) := T(\boldsymbol{x},\boldsymbol{z})\,q(\boldsymbol{z}) = q(\boldsymbol{x}|\boldsymbol{z})\,q(\boldsymbol{z})\}$.

$$
\begin{aligned}
\mathcal{D}(p(\boldsymbol{s})\|q(\boldsymbol{s})) &= \int p(\boldsymbol{s}) \log \frac{p(\boldsymbol{s})}{q(\boldsymbol{s})} d\boldsymbol{s} \\
&= \int p(\boldsymbol{s}) \log p(\boldsymbol{x}) d\boldsymbol{s} + \int p(\boldsymbol{s}) \log \frac{p(\boldsymbol{z}|\boldsymbol{x})}{q(\boldsymbol{z})} d\boldsymbol{s} - \int p(\boldsymbol{s}) \log q(\boldsymbol{x}|\boldsymbol{z}) d\boldsymbol{s} \\
&= -H(\boldsymbol{x}) + \underbrace{R + D}_{\mathcal{L}_{\rm ELBO}} \geq 0
\end{aligned}
$$

VAEの目的関数 変分下限（ELBO）



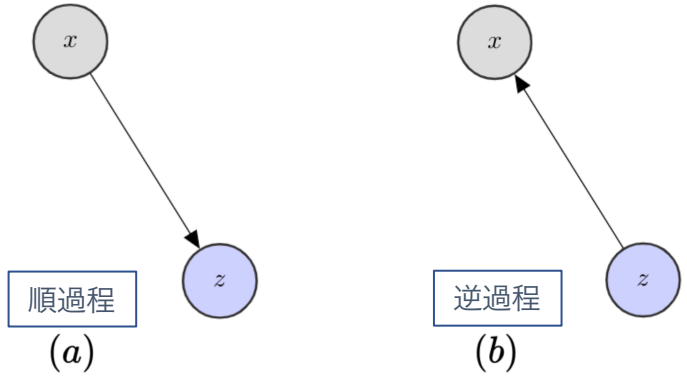
$$
\begin{aligned}
\mathcal{D}(p(\boldsymbol{s})\|q(\boldsymbol{s})) &= \int p(\boldsymbol{s}) \log \frac{p(\boldsymbol{s})}{q(\boldsymbol{s})} d\boldsymbol{s} \\
&= \int p(\boldsymbol{s}) \log \frac{p(\boldsymbol{z}|\boldsymbol{x})p(\boldsymbol{x})}{q(\boldsymbol{x}|\boldsymbol{z})p(\boldsymbol{z})} d\boldsymbol{s} + \int p(\boldsymbol{s}) \log \frac{p(\boldsymbol{z})}{q(\boldsymbol{z})} d\boldsymbol{s} \\
&:= \sigma^{\mathcal{S}}_{\rm bath} + \sigma^{\mathcal{S}}_{\rm sys} = \sigma^{\mathcal{S}}_{\rm tot} \geq 0
\end{aligned}
$$

全系のエントロピー生成率

林祐輔（所属: Japan Digital Design, 連絡先: bheprr@gmail.com, Twitter: @hayashiyus）

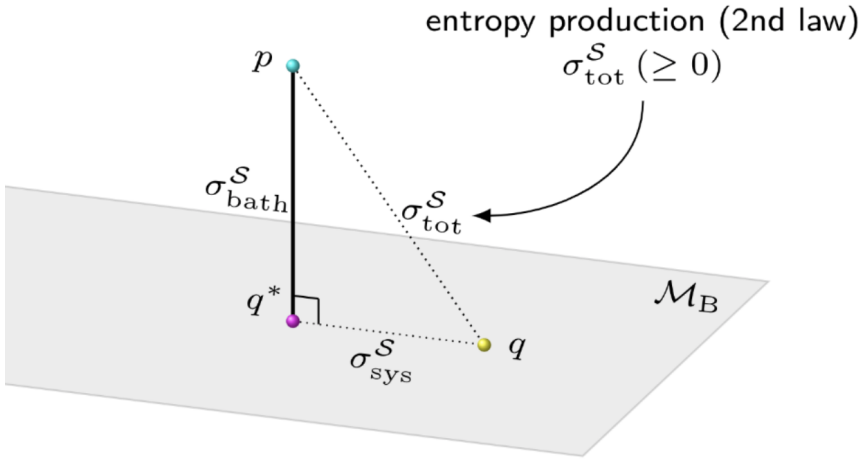
表現学習の熱力学（11/n）

- 全系のエントロピー生成：順過程と逆過程とで，グラフィカルモデルの全要素について矢印の向きが逆になる.



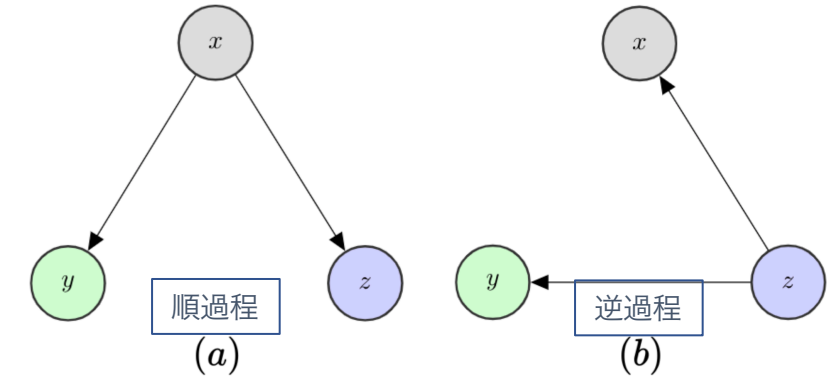
$$p(\boldsymbol{s}) := T(\boldsymbol{z},\boldsymbol{x})\,p(\boldsymbol{x}) = p(\boldsymbol{z}|\boldsymbol{x})\,p(\boldsymbol{x}).$$

$$q(\boldsymbol{s}) := T(\boldsymbol{x},\boldsymbol{z})\,q(\boldsymbol{z}) = q(\boldsymbol{x}|\boldsymbol{z})\,q(\boldsymbol{z})$$



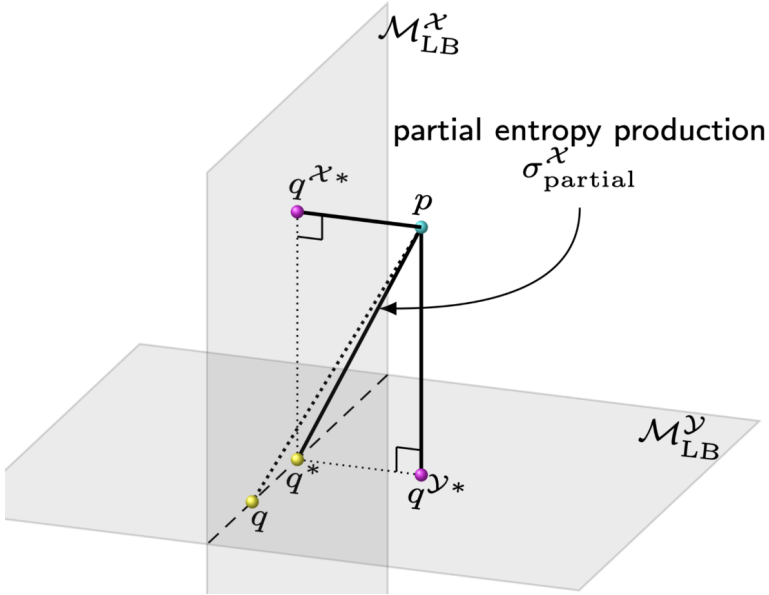
$$\mathcal{M}_{\rm B} = \{q(\boldsymbol{s}) \mid q(\boldsymbol{s}) := T(\boldsymbol{x},\boldsymbol{z})\,q(\boldsymbol{z}) = q(\boldsymbol{x}|\boldsymbol{z})\,q(\boldsymbol{z})\}\,.$$

- 部分系のエントロピー生成：順過程と逆過程とで，グラフィカルモデルの一部の要素について矢印の向きが逆になる.



$$p(\boldsymbol{s}) := p(\boldsymbol{z}|\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})$$

$$q(\boldsymbol{s}) := q(\boldsymbol{x}|\boldsymbol{z})q(\boldsymbol{y}|\boldsymbol{z})\,q(\boldsymbol{z})$$



林祐輔（所属: Japan Digital Design, 連絡先: bheprr@gmail.com, Twitter: @hayashiyus）

表現学習の熱力学（12/n）

- 生成モデルの学習プロセス：データ $\boldsymbol{x}$から潜在因子（特徴量）$\boldsymbol{z}$の生成と教師ラベル $\boldsymbol{y}$の予測を行う順過程の確率分布 pと，潜在因子（特徴量）$\boldsymbol{z}$からデータ $\boldsymbol{x}$と教師ラベル $\boldsymbol{y}$をデコードする逆過程の確率分布 qの距離を可能な限り近づけようとするプロセス.

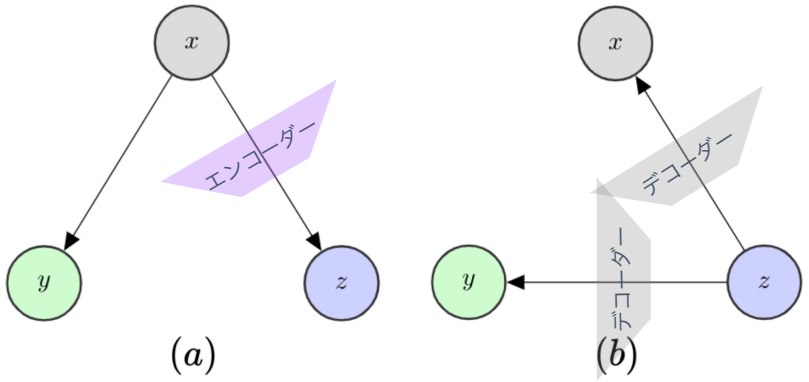
  ⇒ 部分系のエントロピー生成と解釈可能.

$$H := \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}\sim p(\boldsymbol{x},\boldsymbol{y})}\left[-\log p(\boldsymbol{x},\boldsymbol{y})\right], \quad \text{データのエントロピー}$$

$$D := \mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x})}\left[-\int e(\boldsymbol{z}|\boldsymbol{x})\log d(\boldsymbol{x}|\boldsymbol{z})d\boldsymbol{z}\right], \quad \text{再構成誤差}$$

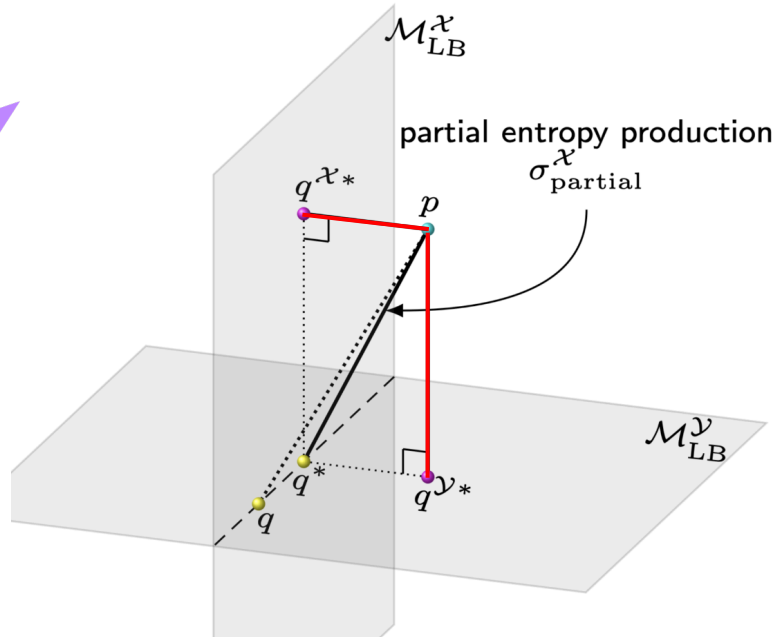
$$R := \mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x})}\left[\int e(\boldsymbol{z}|\boldsymbol{x})\log\frac{e(\boldsymbol{z}|\boldsymbol{x})}{q(\boldsymbol{z})}d\boldsymbol{z}\right], \quad \text{正則化項}$$

$$C := \mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x})}\left[-\int e(\boldsymbol{z}|\boldsymbol{x})\log c(\boldsymbol{y}|\boldsymbol{z})d\boldsymbol{z}\right], \quad \text{分類誤差}$$



$$p(\boldsymbol{s}) := e(\boldsymbol{z}|\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}) \qquad q(\boldsymbol{s}) := d_\theta(\boldsymbol{x}|\boldsymbol{z})^\lambda c_\theta(\boldsymbol{y}_{\boldsymbol{x}}|\boldsymbol{z})^\gamma q_\theta(\boldsymbol{z})$$

$$\mathcal{D}(p(\boldsymbol{s})\|q(\boldsymbol{s})) = -H + R + \lambda D + \gamma C \geq 0$$
$$:= -H + \mathcal{J}(\theta,\lambda,\gamma) \geq 0$$



参考文献[5]: Alemi and Fischer (2018) arXiv:1807.04162v3 **[cs.LG]**

林祐輔（所属: Japan Digital Design, 連絡先: bheprr@gmail.com, Twitter: @hayashiyus）

表現学習の熱力学（13/n）

- 表現学習における熱力学諸法則の導出

$$p(\boldsymbol{s}) := e(\boldsymbol{z}|\boldsymbol{x})p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})$$

$$q(\boldsymbol{s}) := d_\theta(\boldsymbol{x}|\boldsymbol{z})^\lambda c_\theta\left(\boldsymbol{y}_{\boldsymbol{x}}|\boldsymbol{z}\right)^\gamma q_\theta(\boldsymbol{z})$$

$$H := \mathbb{E}_{\boldsymbol{x},\boldsymbol{y}\sim p(\boldsymbol{x},\boldsymbol{y})}\left[-\log p(\boldsymbol{x},\boldsymbol{y})\right], \quad \text{データのエントロピー}$$

$$D := \mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x})}\left[-\int e(\boldsymbol{z}|\boldsymbol{x})\log d(\boldsymbol{x}|\boldsymbol{z})d\boldsymbol{z}\right], \quad \text{再構成誤差}$$

$$R := \mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x})}\left[\int e(\boldsymbol{z}|\boldsymbol{x})\log\frac{e(\boldsymbol{z}|\boldsymbol{x})}{q(\boldsymbol{z})}d\boldsymbol{z}\right], \quad \text{正則化項}$$

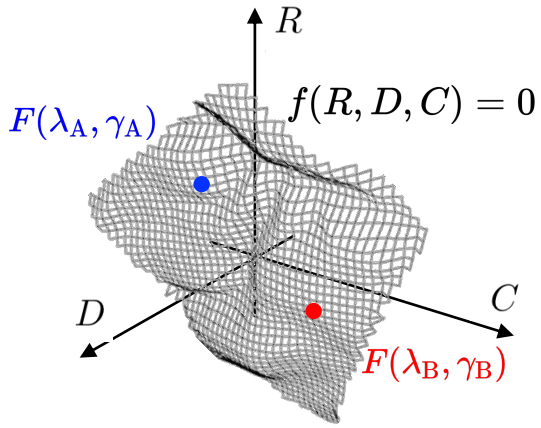
$$C := \mathbb{E}_{\boldsymbol{x}\sim p(\boldsymbol{x})}\left[-\int e(\boldsymbol{z}|\boldsymbol{x})\log c(\boldsymbol{y}|\boldsymbol{z})d\boldsymbol{z}\right], \quad \text{分類誤差}$$

$$\mathcal{D}(p(\boldsymbol{s})\|q(\boldsymbol{s})) = -H + R + \lambda D + \gamma C \geq 0$$
$$:= -H + \mathcal{J}(\theta,\lambda,\gamma) \geq 0$$

$$\mathcal{J}(\theta,\lambda,\gamma) \geq H$$

$$\nabla_\theta \mathcal{J}(\theta^*,\lambda,\gamma) = 0$$
$$f(R,D,C) = 0$$

$$F(\lambda,\gamma) = \min_\theta \mathcal{J}(\theta,\lambda,\gamma)$$
$$= \min_\theta R + \lambda D + \gamma C$$
$$F(\lambda,\gamma) = \mathcal{J}(\theta^*,\lambda,\gamma), \quad \theta^* \in \Theta$$



- 自由エネルギーが定義される最適曲面の上では，以下が成り立つ．

Remark 1 (“the first law” of learning).

$$\mathrm{d}R = -\lambda \mathrm{d}D - \gamma \mathrm{d}C \qquad \lambda = -\left(\frac{\partial R}{\partial D}\right)_C \quad \gamma = -\left(\frac{\partial R}{\partial C}\right)_D \quad D = \left(\frac{\partial F}{\partial \lambda}\right)_\gamma \quad C = \left(\frac{\partial F}{\partial \gamma}\right)_\lambda$$

Remark 2 (“the second law” of learning).

$$\mathcal{D}(p(\boldsymbol{s})\|q(\boldsymbol{s})) = -H + R + \lambda D + \gamma C := -H + \mathcal{J}(\theta,\lambda,\gamma) \geq 0$$

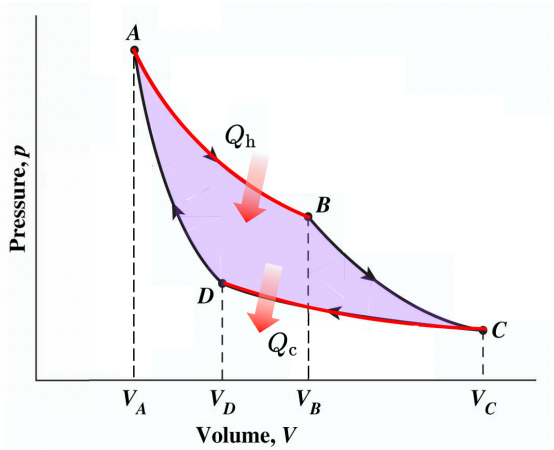
参考文献[5]: Alemi and Fischer (2018) arXiv:1807.04162v3 **[cs.LG]**

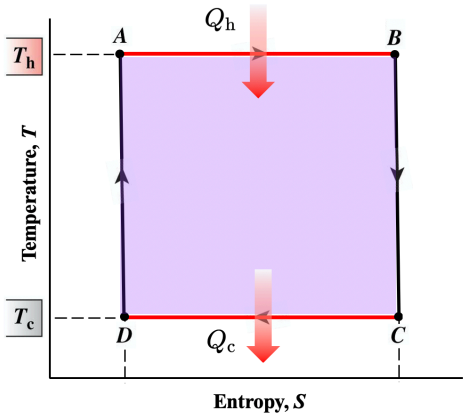
林祐輔（所属: Japan Digital Design, 連絡先: bheprr@gmail.com, Twitter: @hayashiyus）

表現学習の熱力学（14/n）

- 等温準静過程のアナロジーを考えることで，統計モデルの分類誤差を低い値に抑えたまま
  データのドメインを変化させること（＝転移学習）はできないか？

熱力学

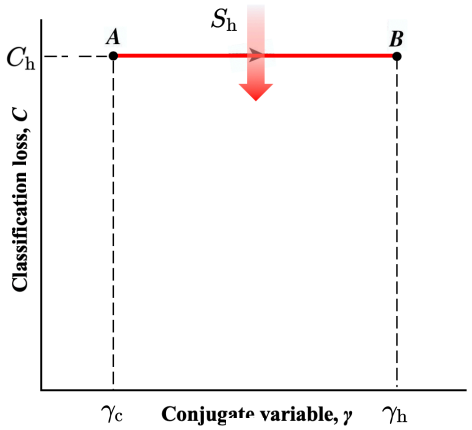




$$F(\lambda,\gamma) = \min_{\theta} R + \lambda D + \gamma C$$

$$F = U - TS$$
$$\Delta U = Q + W$$
$$F_{\mathrm{B}} - F_{\mathrm{A}} = W_{\mathrm{AB}}$$
$$Q_{\mathrm{AB}} = U_{\mathrm{B}} - U_{\mathrm{A}} - (F_{\mathrm{B}} - F_{\mathrm{A}})$$

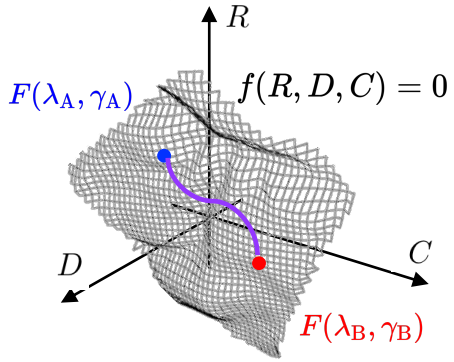
表現学習

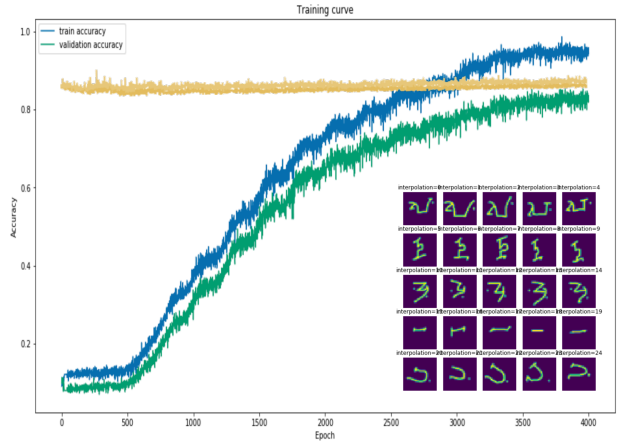


- データのドメインを変化させる間，分類誤差の値が一定となるような“熱力学的操作”を考える.

$$p(x(t)) = \underset{p(x)}{\operatorname{argmin}} (1-t) W_2^2\left(p(x^{\mathrm{A}}), p(x)\right) + t W_2^2\left(p(x), p(x^{\mathrm{B}})\right)$$

$$\frac{\mathrm{d}C(t)}{\mathrm{d}t} = 0$$





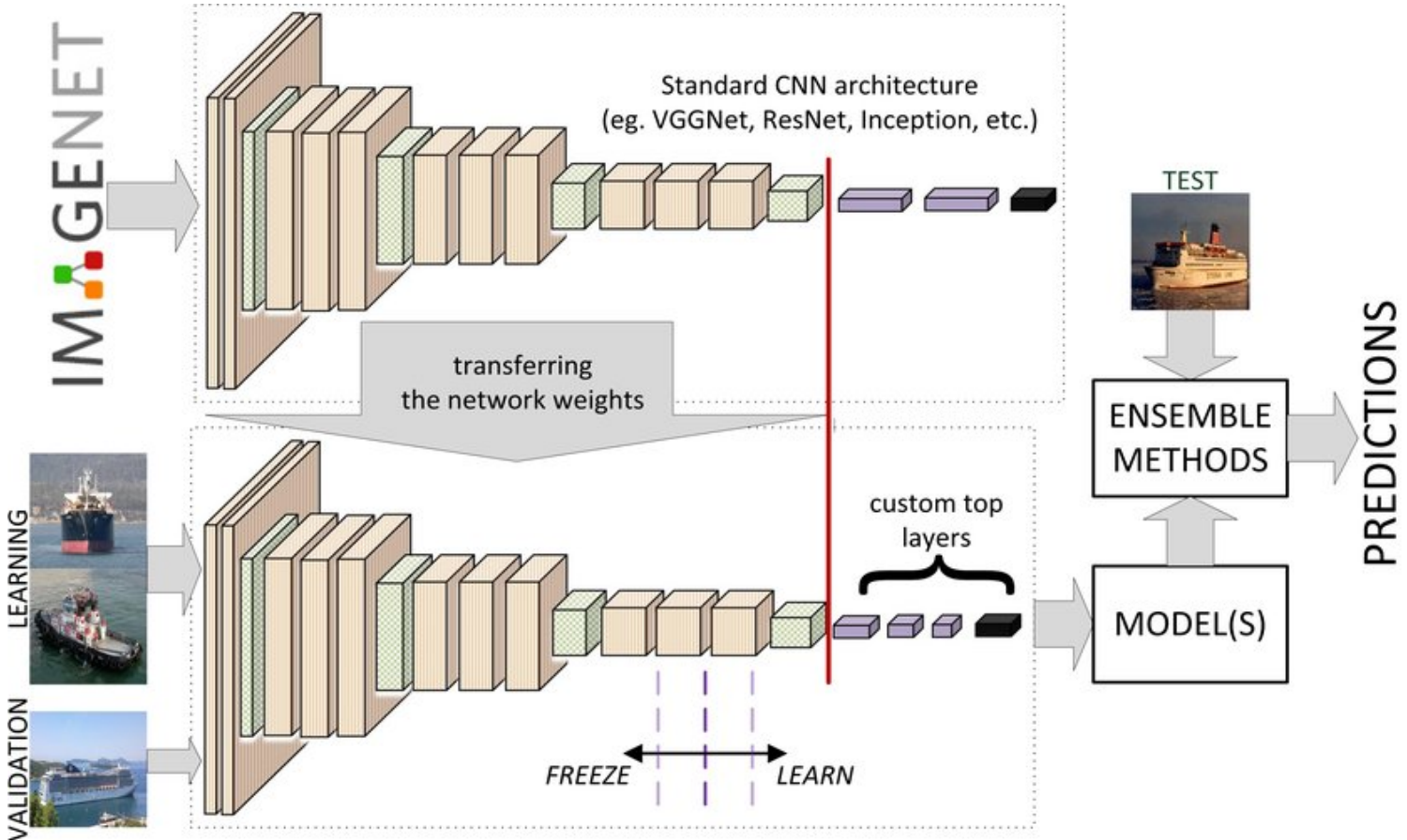
MNIST の 0〜4 までの数字（ドメイン）について訓練
⇒ 残った 5〜9 までの数字（ドメイン）について転移学習

参考文献[6]: Gao and Chaudhari (2020) arXiv:2002.12406v1 **[cs.LG]**

林祐輔（所属: Japan Digital Design, 連絡先: bheprr@gmail.com, Twitter: @hayashiyus）

転移学習とはなにか（15/n）

特定のドメインで訓練が終了した学習済みモデルを，別のドメインの学習に再利用すること．少ない訓練データで高い分類性能を引き出すアルゴリズムを考えたりすることができる．



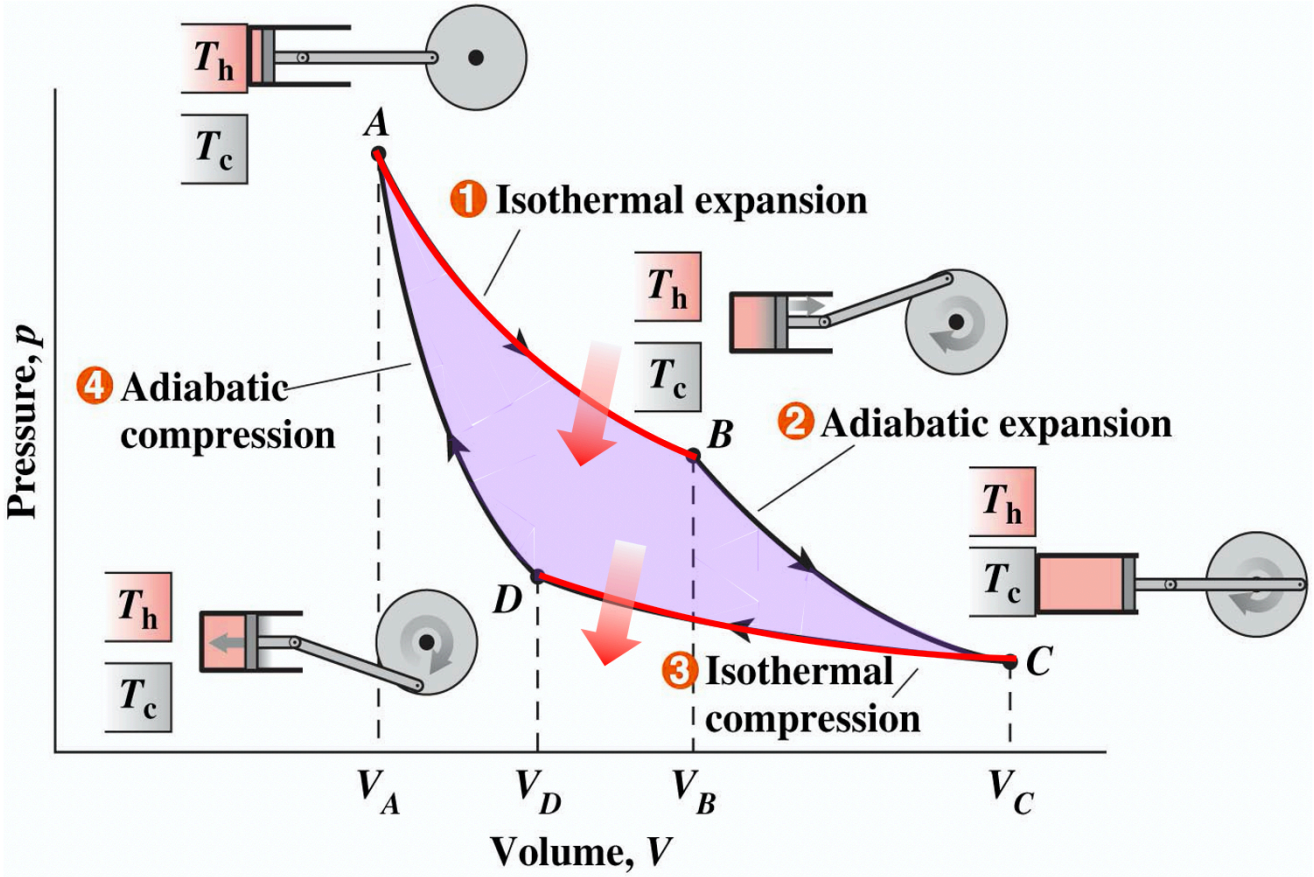
https://medium.com/@subodh.malgonde/transfer-learning-using-tensorflow-52a4f6bcde3e

学習済みモデルのパラメータを，別のドメインのデータの学習に再利用

林祐輔（所属: Japan Digital Design, 連絡先: bheprr@gmail.com, Twitter: @hayashiyus）

熱力学エンジンの拘束条件（16/n）

自由エネルギーや準静過程の意味は，熱力学エンジンの最大熱効率を求めるカルノーの定理を考えるとはっきりしてくる．カルノーサイクルに関する議論を振り返る．



① 等温準静過程（膨張）

高温熱浴に接したまま膨張

② 断熱準静過程（膨張）

エンジンを熱浴から離して断熱膨張

③ 等温準静過程（圧縮）

低温熱浴に接したまま圧縮

④ 断熱準静過程（圧縮）

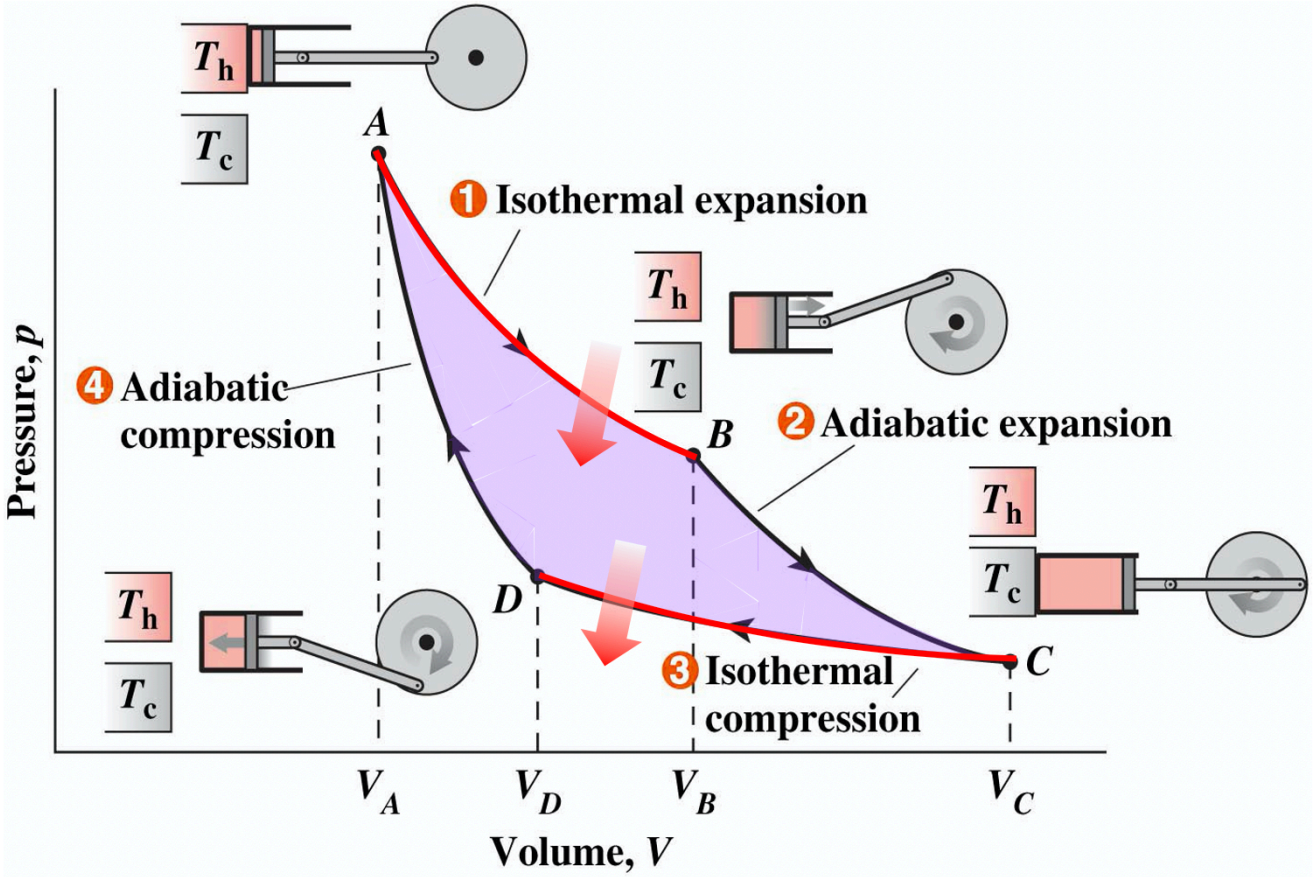
エンジンを熱浴から離して断熱圧縮

林祐輔（所属: Japan Digital Design, 連絡先: bheprr@gmail.com, Twitter: @hayashiyus）

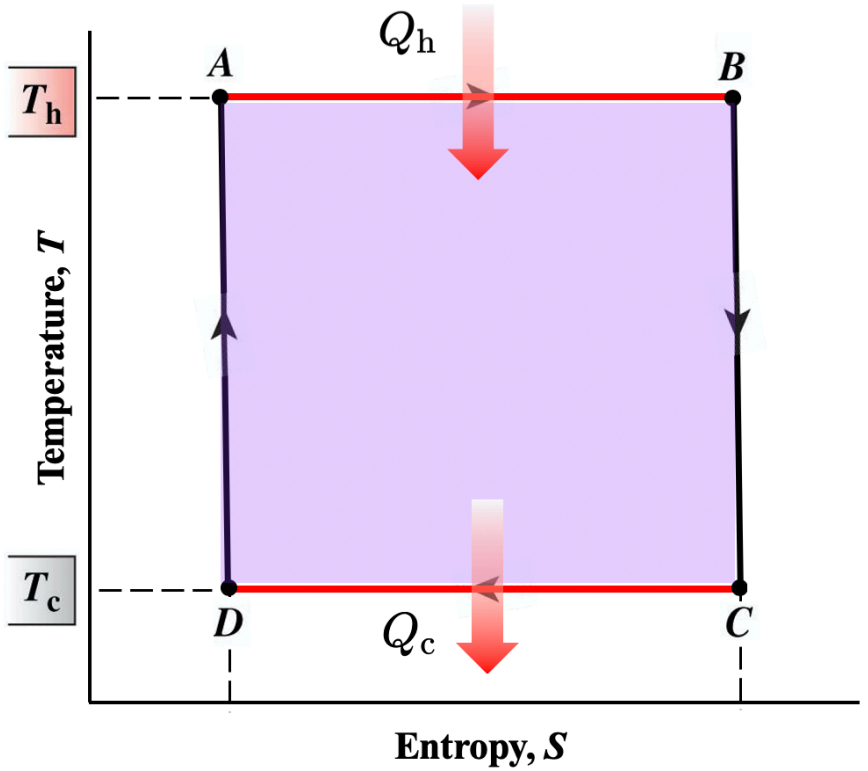
熱力学エンジンの拘束条件（17/n）

- pV-平面上に描かれた準静サイクルを，TS-平面上に描き直すと平行四辺形があらわれる.

pV-平面



TS-平面

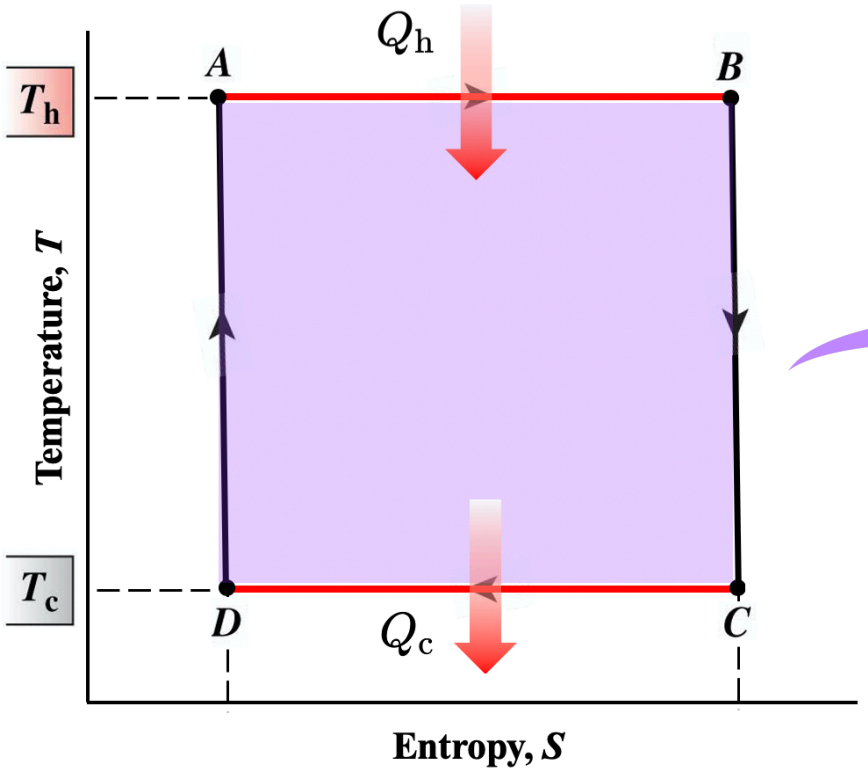


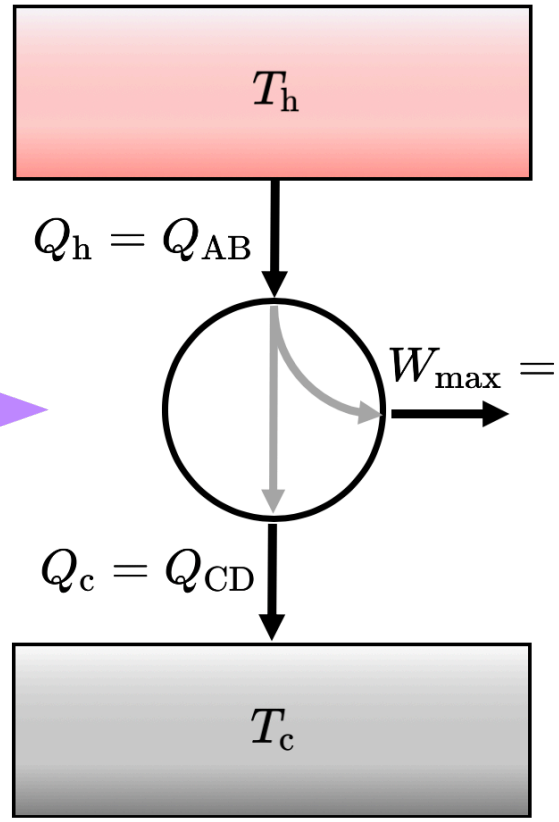
林祐輔（所属: Japan Digital Design, 連絡先: bheprr@gmail.com, Twitter: @hayashiyus）

熱力学エンジンの拘束条件（18/n）

- カルノーの定理より，準静過程を組み合わせた準静サイクルが最大熱効率（カルノー効率）を与えることが示される．

TS-平面





カルノー効率

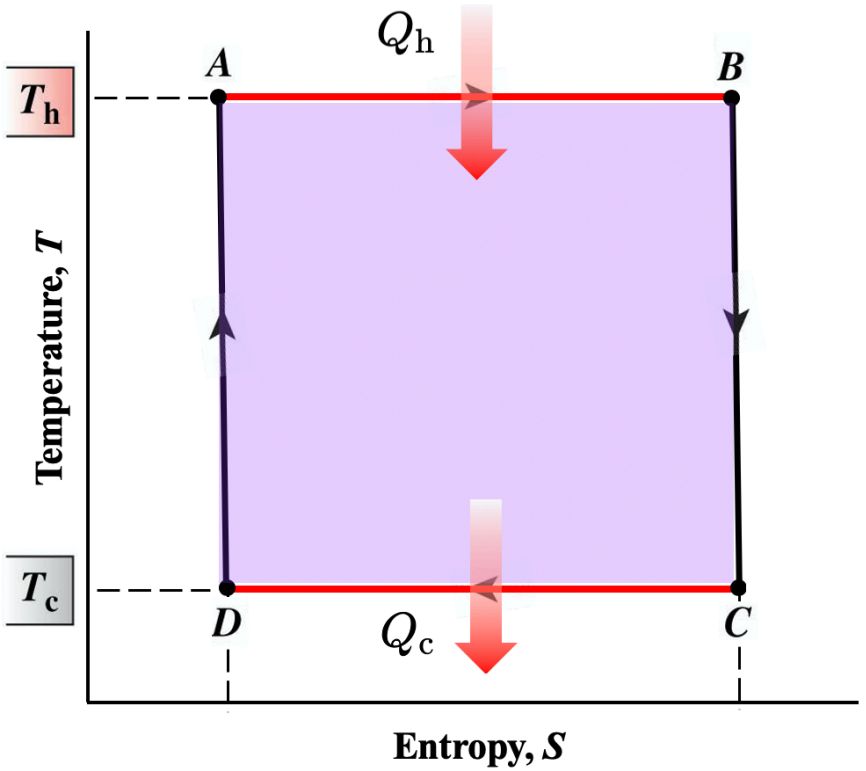
$$\eta_{\mathrm{C}} := \frac{W_{\max}}{Q_{\mathrm{h}}} = 1 - \frac{Q_{\mathrm{c}}}{Q_{\mathrm{h}}}$$

林祐輔（所属: Japan Digital Design, 連絡先: bheprr@gmail.com, Twitter: @hayashiyus）

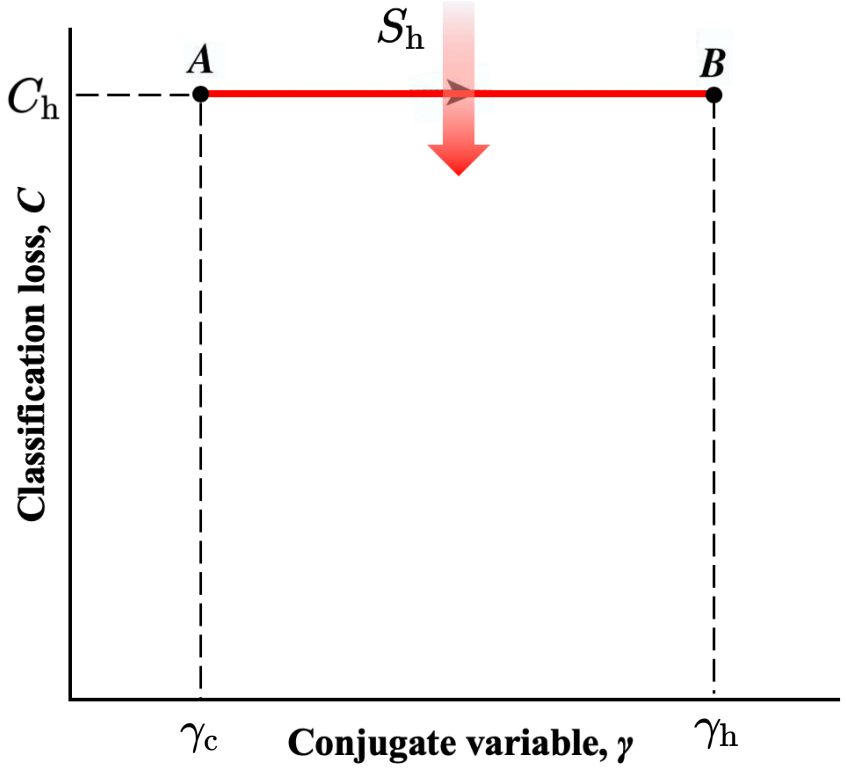
熱力学エンジンの拘束条件（19/n）

- 等温準静過程のアナロジーを考えることで，統計モデルの分類誤差を低い値に抑えたままデータのドメインを変化させること（＝転移学習）はできないか？

熱力学



表現学習



参考文献[6]: Gao and Chaudhari (2020) arXiv:2002.12406v1 **[cs.LG]**

林祐輔（所属: Japan Digital Design, 連絡先: bheprr@gmail.com, Twitter: @hayashiyus）

表現学習の熱力学（20/n）

- 分類誤差を“温度”にみたてて，等温準静過程に沿ってモデルのパラメータ等を動かしていく．

$$p(x(t)) = \underset{p(x)}{\operatorname{argmin}} (1-t) W_2^2\left(p(x^{\mathrm{A}}), p(x)\right) + t W_2^2\left(p(x), p(x^{\mathrm{B}})\right)$$

要請①

$$C(t) := \mathbb{E}_{\boldsymbol{x}(t)\sim p(\boldsymbol{x}(t))}\left[-\int e(\boldsymbol{z}(t) \mid \boldsymbol{x}(t)) \log c(\boldsymbol{y}(t) \mid \boldsymbol{z}(t)) d\boldsymbol{z}(t)\right]$$

$$\frac{\mathrm{d}C(t)}{\mathrm{d}t} = \dot{\theta}^\top \nabla_\theta C + \dot{\lambda}\frac{\partial C}{\partial \lambda} + \dot{\gamma}\frac{\partial C}{\partial \gamma} = 0$$

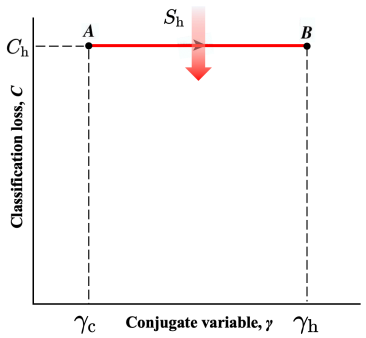
要請②

$$\nabla_\theta \mathcal{J}(\theta^*(t), \lambda(t), \gamma(t)) = 0$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\nabla_\theta \mathcal{J}(\theta^*, \lambda, \gamma) = \dot{\theta}^\top \nabla_\theta^2 \mathcal{J}_{\theta^*} + \dot{\lambda}\frac{\partial}{\partial \lambda}\nabla_\theta \mathcal{J}_{\theta^*} + \dot{\gamma}\frac{\partial}{\partial \gamma}\nabla_\theta \mathcal{J}_{\theta^*} = 0$$

要請③

$$\dot{\lambda} = k$$



$$\dot{\theta}^\top \nabla_\theta C + \dot{\lambda}\frac{\partial C}{\partial \lambda} + \dot{\gamma}\frac{\partial C}{\partial \gamma} = 0$$

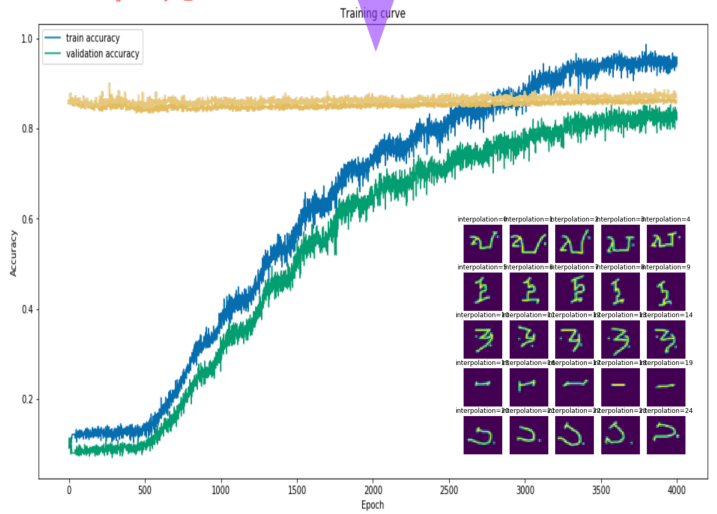
$$\dot{\theta}^\top \nabla_\theta^2 \mathcal{J} + \dot{\lambda}\frac{\partial}{\partial \lambda}\nabla_\theta \mathcal{J} + \dot{\gamma}\frac{\partial}{\partial \gamma}\nabla_\theta \mathcal{J} = 0$$

$$\dot{\lambda} = k$$

$(\theta, \lambda, \gamma)$ についての連立一次確率微分方程式

どういうときに転移学習が成功して

どういうときには失敗するのか不明



MNIST の 0〜4 までの数字（ドメイン）について訓練
⇒ 残った 5〜9 までの数字（ドメイン）について転移学習

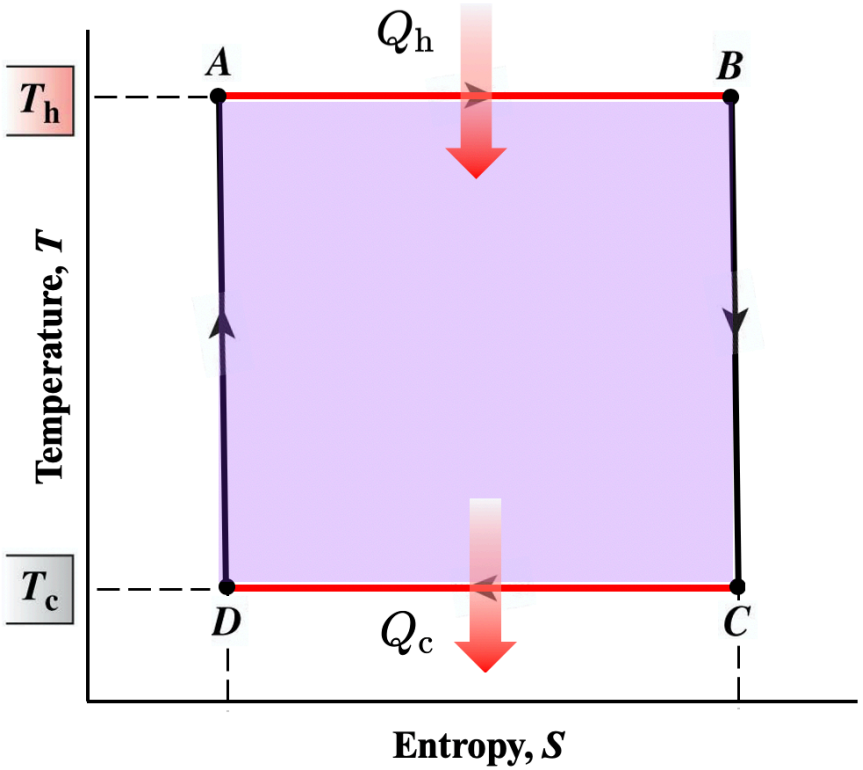
参考文献[6]: Gao and Chaudhari (2020) arXiv:2002.12406v1 [cs.LG]

林祐輔（所属: Japan Digital Design, 連絡先: bheprr@gmail.com, Twitter: @hayashiyus）

表現学習の熱力学（21/n）

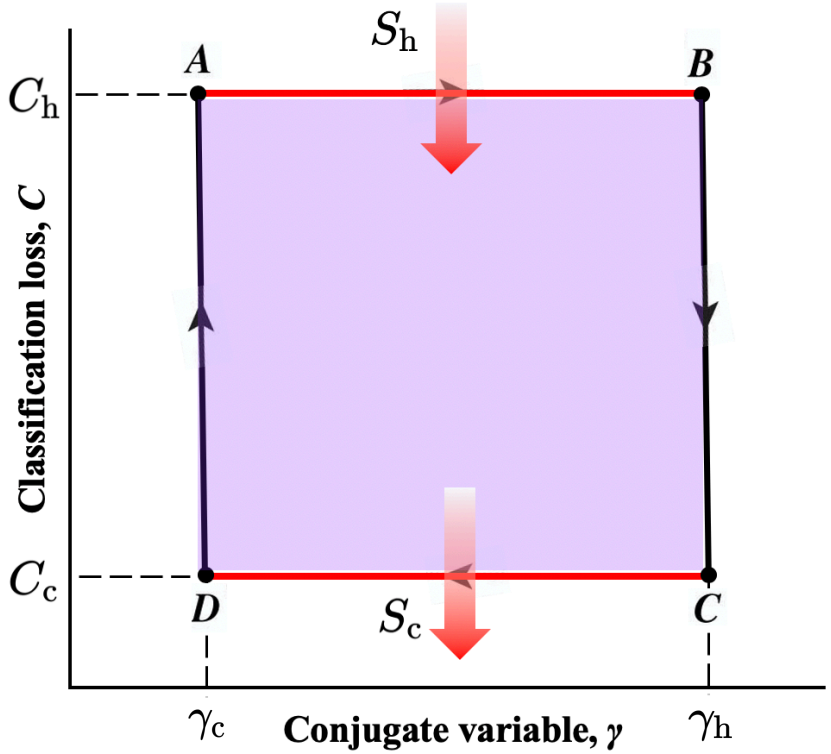
- 等温準静過程に加え，断熱準静過程に対応する熱力学的操作も導入することができれば表現学習版のカルノーサイクルを考えることができる.

熱力学



$$\eta_C := \frac{W_{max}}{Q_h} = 1 - \frac{Q_c}{Q_h}$$

表現学習



$$\eta_C := \frac{W_{max}}{S_h} = 1 - \frac{S_c}{S_h}$$

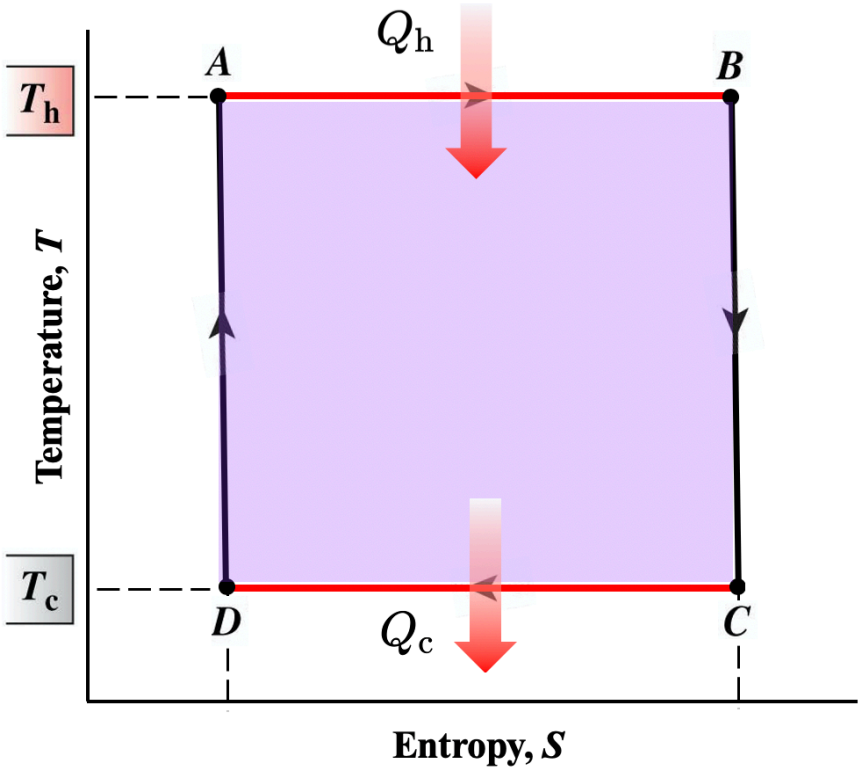
?

林祐輔（所属: Japan Digital Design, 連絡先: bheprr@gmail.com, Twitter: @hayashiyus）

表現学習の熱力学（22/n）

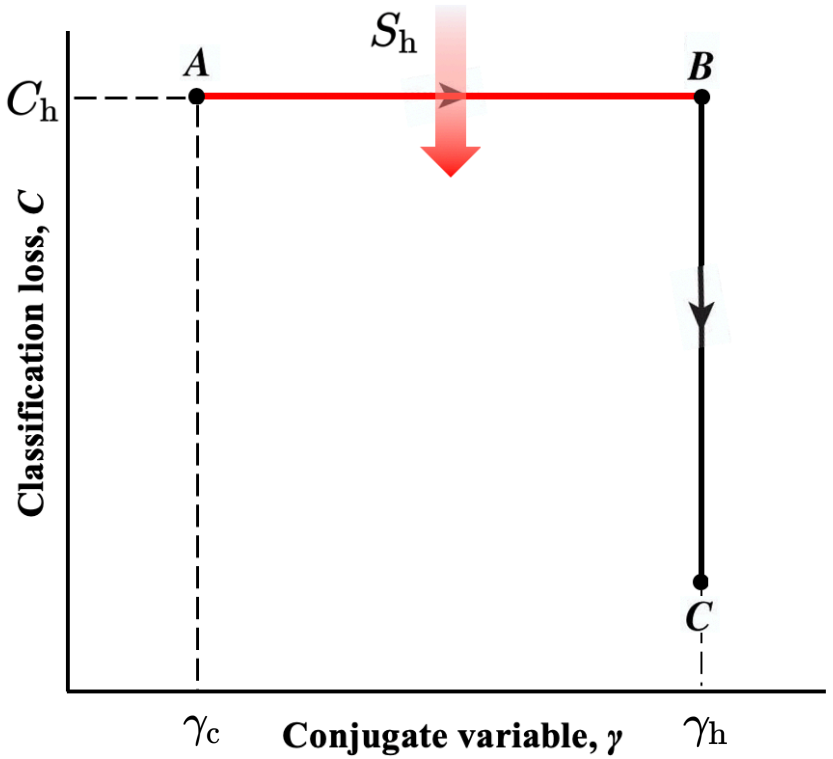
- 表現学習版のカルノーサイクルを考えるために，断熱準静過程に対応する熱力学的操作を導入したい．

熱力学



$$\eta_C := \frac{W_{max}}{Q_h} = 1 - \frac{Q_c}{Q_h}$$

表現学習



$$\eta_C := \frac{W_{max}}{S_h} = 1 - \frac{S_c}{S_h}$$ ？

林祐輔（所属: Japan Digital Design, 連絡先: bheprr@gmail.com, Twitter: @hayashiyus）

表現学習の熱力学（23/n）

- 断熱準静過程のアナロジーを表現学習に持ち込む.

$$p(x(t)) = \underset{\mathrm{p(x)}}{\operatorname{argmin}} (1-t) W_2^2\left(p\left(x^{\mathrm{B}}\right), p(x)\right) + t W_2^2\left(p(x), p\left(x^{\mathrm{C}}\right)\right)$$

要請①

$$\gamma = -\left(\frac{\partial R}{\partial C}\right)_D$$

$$\dot{\gamma} = \alpha \frac{\partial C}{\partial \lambda} = \alpha \frac{\partial^2 F}{\partial \lambda \partial \gamma} = \alpha \frac{\partial^2 \mathcal{J}_{\theta^*}}{\partial \lambda \partial \gamma} = 0$$

要請②

$$\nabla_\theta \mathcal{J}(\theta^*(t), \lambda(t), \gamma(t)) = 0$$

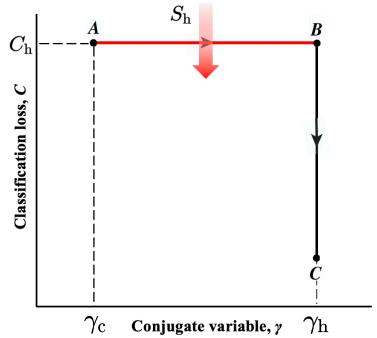
$$\frac{\mathrm{d}}{\mathrm{d}t} \nabla_\theta \mathcal{J}(\theta^*, \lambda, \gamma) = \dot{\theta}^\top \nabla_\theta^2 \mathcal{J}_{\theta^*} + \dot{\lambda} \frac{\partial}{\partial \lambda} \nabla_\theta \mathcal{J}_{\theta^*} + \dot{\gamma} \frac{\partial}{\partial \gamma} \nabla_\theta \mathcal{J}_{\theta^*} = 0$$

要請③

$$\dot{\lambda} = k$$

要請④（断熱条件）

$$S_{\mathrm{BC}} = R_{\mathrm{C}} - R_{\mathrm{B}} - (F_{\mathrm{C}} - F_{\mathrm{B}}) = 0$$



$$\dot{\gamma} = \alpha \frac{\partial^2 \mathcal{J}}{\partial \lambda \partial \gamma} = 0$$

$$\dot{\theta}^\top \nabla_\theta^2 \mathcal{J} + \dot{\lambda} \frac{\partial}{\partial \lambda} \nabla_\theta \mathcal{J} + \dot{\gamma} \frac{\partial}{\partial \gamma} \nabla_\theta \mathcal{J} = 0$$

$$\dot{\lambda} = k$$

$$S_{\mathrm{BC}} = R_{\mathrm{C}} - R_{\mathrm{B}} - (F_{\mathrm{C}} - F_{\mathrm{B}}) = 0$$

$(\theta, \lambda, \gamma)$ についての連立一次確率微分方程式

$$S_{\mathrm{AB}} = R_{\mathrm{B}} - R_{\mathrm{A}} - (F_{\mathrm{B}} - R_{\mathrm{A}})$$
$$= -\int_{t_{\mathrm{A}}}^{t_{\mathrm{B}}} \lambda(t) D(t) + \gamma(t) C(t) dt$$

$$\int_{t_{\mathrm{A}}}^{t_{\mathrm{B}}} \gamma(t) C(t) dt = C \int_{t_{\mathrm{A}}}^{t_{\mathrm{B}}} \gamma(t) dt$$

$$p(x^{\mathrm{A}}) \longrightarrow p(x^{\mathrm{B}})$$

林祐輔（所属: Japan Digital Design, 連絡先: bheprr@gmail.com, Twitter: @hayashiyus）

表現学習の熱力学（24/n）

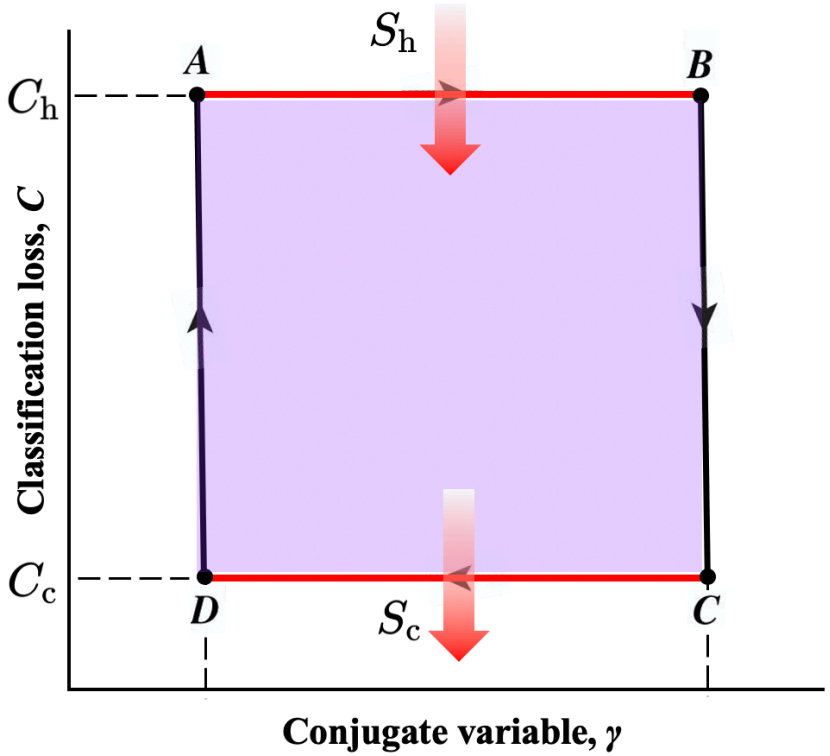
- 表現学習版の熱力学第1法則・第2法則から，カルノーサイクルが最大熱効率を与えることを確認.

Remark 1 (“the first law” of learning).

$$\mathrm{d}R = -\lambda \mathrm{d}D - \gamma \mathrm{d}C \qquad \lambda = -\left(\frac{\partial R}{\partial D}\right)_C \quad \gamma = -\left(\frac{\partial R}{\partial C}\right)_D \quad D = \left(\frac{\partial F}{\partial \lambda}\right)_\gamma \quad C = \left(\frac{\partial F}{\partial \gamma}\right)_\lambda$$

Remark 2 (“the second law” of learning).

$$\mathcal{D}(p(s)\|q(s)) = -H + R + \lambda D + \gamma C := -H + \mathcal{J}(\theta, \lambda, \gamma) \geq 0$$



最大熱効率（カルノー効率）

$$\eta_\mathrm{C} := \frac{W_\mathrm{max}}{S_\mathrm{h}} = 1 - \frac{S_\mathrm{c}}{S_\mathrm{h}}$$

参考文献[7]: Hideaki Shimazaki, Neural Engine Hypothesis (Dynamic Neuroscience pp267-291), Springer (2017)

林祐輔（所属: Japan Digital Design, 連絡先: bheprr@gmail.com, Twitter: @hayashiyus）

表現学習の熱力学（25/n）

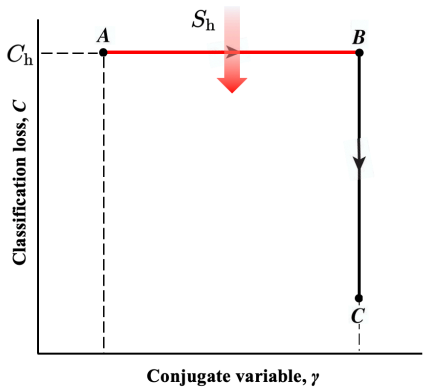
- まとめ
  - エントロピー生成と生成モデルの損失関数を同一視することで，変分下限と自由エネルギーを対応させる関係式を発見.

$$\sigma_{\mathrm{tot}}^{\mathcal{S}} = \beta[W - \Delta F] \qquad \sigma_{\mathrm{tot}}^{S} = -H + \underbrace{R + \lambda D + \gamma C}_{\mathcal{L}_{\mathrm{ELBO}}}$$

  - “断熱準静過程”を表現学習に導入.

ドメインBからCへの転移

$$p(x(t)) = \underset{\mathrm{p(x)}}{\operatorname{argmin}}(1-t)W_2^2\left(p\left(x^{\mathrm{B}}\right), p(x)\right) + tW_2^2\left(p(x), p\left(x^{\mathrm{C}}\right)\right)$$



  - 表現学習版のカルノーサイクルを考え，これが最大熱効率を与えることを確認.

$$\eta_{\mathrm{C}} := \frac{W_{\max}}{S_{\mathrm{h}}} = 1 - \frac{S_{\mathrm{c}}}{S_{\mathrm{h}}}$$

林祐輔（所属: Japan Digital Design, 連絡先: bheprr@gmail.com, Twitter: @hayashiyus）

Appendix（26/n）

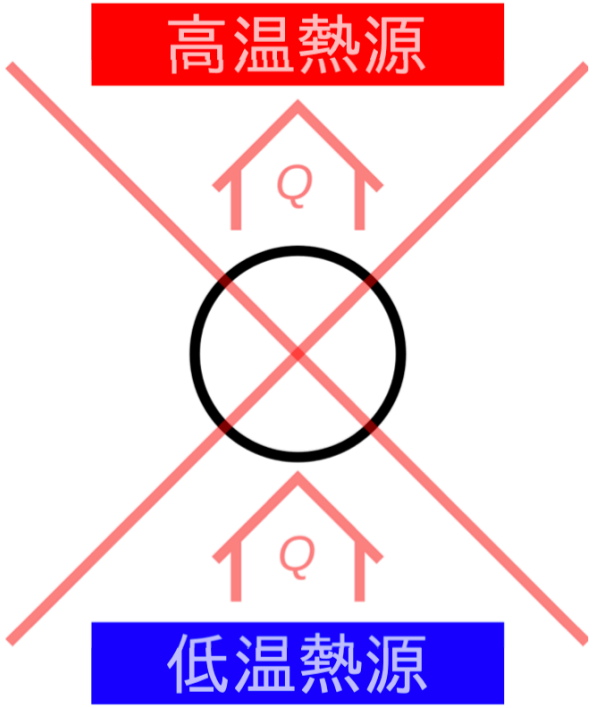


図 1.3　クラウジウスの原理

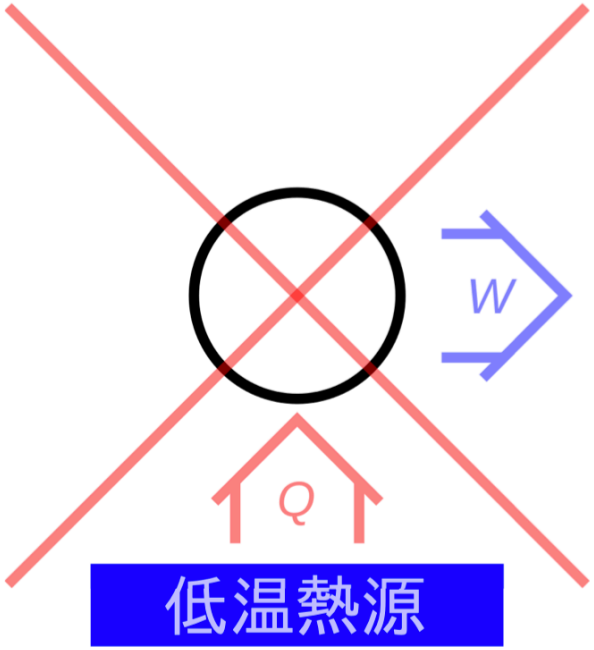


図 1.4　トムソンの原理

https://camelllia.net/study.html