

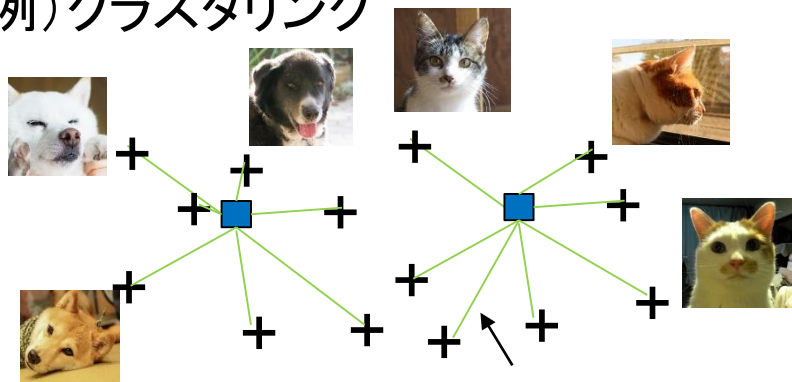
レート歪み関数と最適再構成分布

豊橋技術科学大学 情報・知能工学系 渡辺一帆

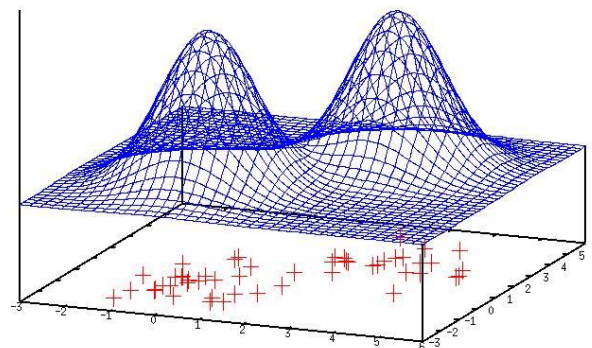
機械学習とレート歪み理論(1)

- 機械学習

例) クラスタリング



$d(x, y)$: 損失関数



混合正規分布: 潜在変数モデル

- 学習モデルの設計

損失関数, モデルの複雑さ, 事前分布(正則化項)

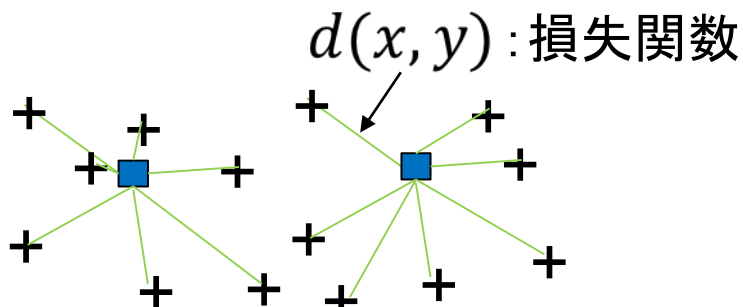
- 潜在変数モデルの学習: レート歪み理論的な解釈

[Watanabe, Ikeda, 2015]

機械学習とレート歪み理論(2)

歪み有りデータ圧縮

+ : データ ■ : クラスタ中心 (=再構成)



レート: $\log(\text{クラスタ数})/\text{次元}$

歪み: 各データの損失の平均

□ 学習法の性能解析や解釈

分類 [Nokleby et al., 2016]

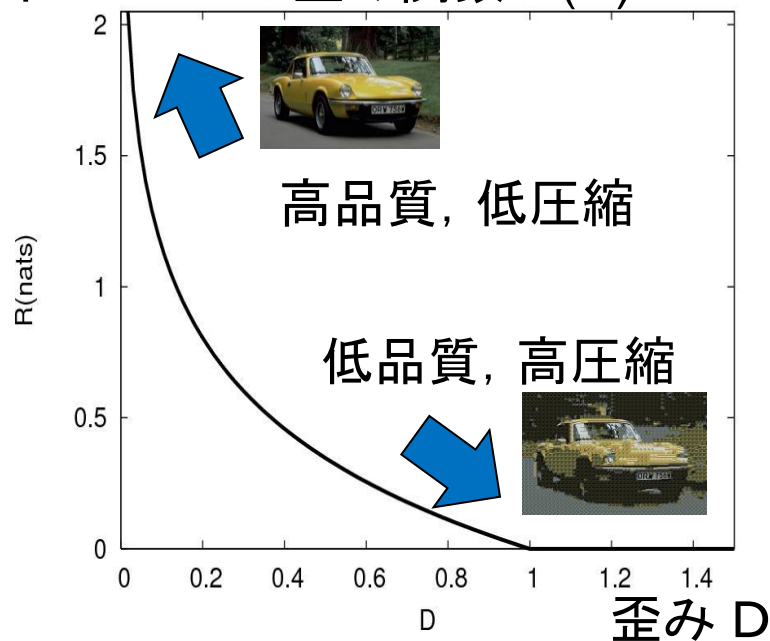
系列予測 [Weissman, Merhav, 2003]

深層学習関連: 情報ボトルネック [Tishby et al., 1999]

オートエンコーダ [Kingma, Welling, 2013]

相互情報量による汎化誤差解析 [Xu, Raginsky, 2017] ...

レート R レート歪み関数 $R(D)$



レート歪み関数

- レート歪み関数(定義)

$$R(D) = \inf_{q(y|x): E[d(X,Y)] \leq D} I(q)$$

$p(x)$: 情報源

$d(x, y)$: 歪み尺度

$I(q) = I(X; Y) = E_{p(x)}[\text{KL}(q(y|X) || q(y))]$: 相互情報量

$q(y) = \int q(y|x)p(x)dx$: (周辺)再構成分布

- 操作的意味

(ϕ_n, ψ_n) : 符号器-復号器の対

$$Y^n = \psi_n(\phi_n(X^n))$$

$\{1, 2, \dots, 2^{nR}\}$ nR bits

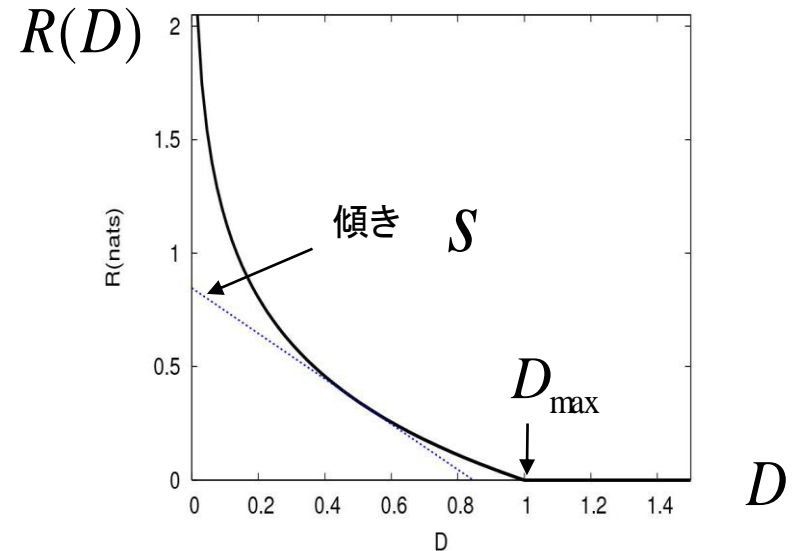
$$\frac{1}{n} \sum_{i=1}^n d(X_i, Y_i)$$

$$\lim_{n \rightarrow \infty} E[d(X^n, Y^n)] \leq D$$

$R(D)$ は達成可能なレート R の下限。

R(D)の性質

- 単調非増加
- 凸
- $D_{\max} = \inf_y \int d(x, y) p(x) dx$
- 傾きパラメータ $s \leq 0$



○ パラメータ表示

← ラグランジュ未定乗数

$$R(D_s) = \min_{q(y|x)} \{I(q) - sE[d(X, Y)]\} + sD_s$$

$$D_s = \int p(x) \underline{q_s(y|x)} d(x, y) dx dy$$

← 最適解

二重最小化としての見方

$$\begin{aligned}
 R(D_s) &= \min_{q(y|x)} \{I(q) - sE[d(X, Y)]\} + sD_s \\
 &= \min_{q(y|x)} \int q(y|x)p(x) \log \frac{q(y|x)}{\int q(y|x')p(x')dx' \exp(sd(x, y))} dx dy + sD_s \\
 &= \min_{q(y|x)} \min_{q(y)} \int q(y|x)p(x) \log \frac{q(y|x)}{q(y) \exp(sd(x, y))} dx dy + sD_s
 \end{aligned}$$

∴ 最小は $q(y) = \int q(y|x')p(x')dx'$ (*)

$$= \min_{q(y)} E_{p(x)} \left[-\log \int \exp(sd(X, y))q(y)dy \right] + sD_s$$

∴ 最小は $q(y|x) \propto q(y) \exp(sd(x, y))$ (+)

(*)と(+)
の繰り返し: Arimoto-Blahutのアルゴリズム

(q(y)の台が有限個の点で既知の場合)

最適再構成分布

自由エネルギー(汎)関数

$$\min_{q(y)} E_{p(x)} \left[-\log \int \exp(sd(X, y)) q(y) dy \right]$$



シャノンの最適再構成分布

最小化する $q(y)$

KKT条件

$$\begin{cases} c(y) = 1 & (y \in V_s) \\ c(y) \leq 1 & (y \notin V_s) \end{cases}$$

$V_s = \{y; q(y) > 0\}$: 再構成分布の台

$$c(y) \equiv \int \frac{p(x) \exp(sd(x, y))}{\int q(\tilde{y}) \exp(sd(x, \tilde{y})) d\tilde{y}} dx$$

シャノンの下界

- 差分歪み尺度 $d(x, y) = \rho(x - y)$

$$g_s(z) \propto e^{s\rho(z)} \quad (g_s * q)(x) = \int g_s(x - y)q(y)dy$$

(無限次元)混合族へのm-射影

$$R(D) = \max_{s \leq 0} \left[\min_{q(y)} \underbrace{KL(p \parallel g_s * q)}_{\text{微分エントロピー}} + h(p) + sD - \log \int e^{s\rho(z)} dz \right]$$

$$\geq \max_{s \leq 0} \left\{ h(p) + sD - \log \int e^{s\rho(z)} dz \right\}$$

$$= h(p) - h(g_{s_L}) \quad \int \rho(x) g_{s_L}(x) dx = D$$

$$\equiv R_L(D) \quad \text{: Shannon lower bound (SLB)}$$

- $R(D) = R_L(D) \iff p = g_s * q \iff Q(\omega) = \frac{P(\omega)}{G_S(\omega)}$: 特性関数

レート歪み関数の解析

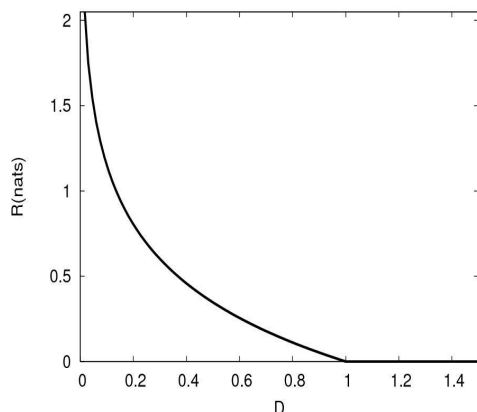
- レート歪み関数 $R(D)$ が解明されている例

損失関数: $d(x, y)$	情報源: $p(x)$	最適再構成分布: $q(y)$	SLB	文献
二乗損失 $(x - y)^2$	ガウス	ガウス	=	[Shannon, 1948]
	ガウス以外	離散	>	[Fix, 1978] [Rose, 1994]
絶対損失 $ x - y $	ラプラス	離散+ラプラス	=	[Berger, 1971] [Verdu, 1996]
	ガウス, 一様等	離散+連続	>	[Tan, Yao, 1975] [Yao, Tan, 1978]
板倉・斉藤損失 $y - x + e^{x-y} - 1$	ガンマ	ベータ	=	[Buzo et al., 1986]

レート歪み関数の解析例

• $p(x) = N(x; \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{x^2}{2\sigma^2}\right)$: ガウス分布

$$d(x, y) = (x - y)^2$$



⇔ $q(y) = N(y; \sigma^2 - D)$
 $R(D) = \frac{1}{2} \log \frac{\sigma^2}{D} \quad (0 \leq D \leq \sigma^2)$

• $p(x) = l(x; \alpha) = \frac{\alpha}{2} e^{-\alpha|x|}$: ラプラス分布

$$d(x, y) = |x - y|$$

⇔ $q(y) = \alpha^2 D^2 \delta(y) + (1 - \alpha^2 D^2) l(y; \alpha)$
 $R(D) = \log \frac{1}{\alpha D} \quad (0 \leq D \leq 1/\alpha)$

Diracのデルタ関数

べき乗損失

$$d(x, y) = |x - y|^\beta \quad p(x) \propto \exp(-|\alpha x|^{\beta'})$$

- $R_L(D) = R(D) (\forall D)$ となるのは, $\beta = \beta'$ のとき
- β にずれがあっても, 近い上界を作れる
[Watanabe, 2017]

対数凹情報源一般の場合:

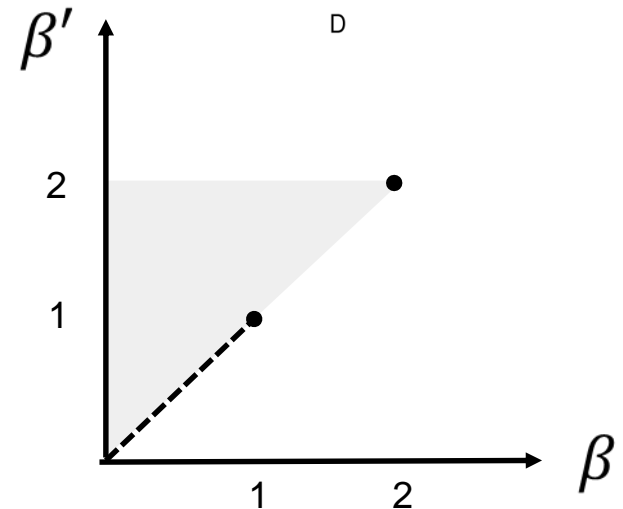
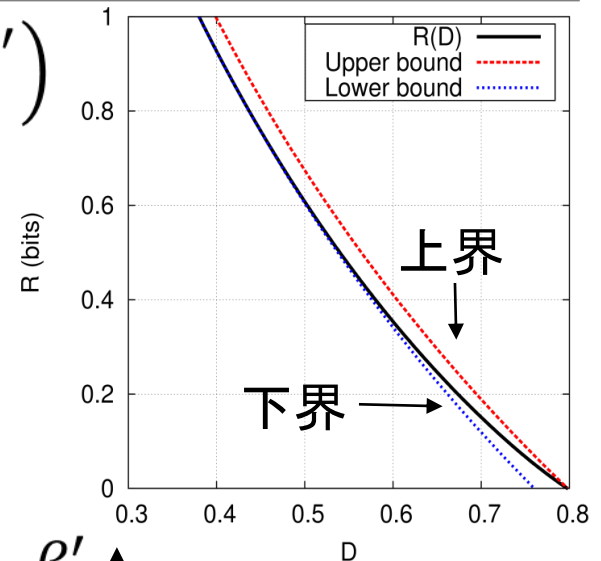
[Marsiglietti, Kostina, 2018]

- $\beta \in (0, 1] \cup \{2\}$ では, $R_L(D) = R(D)$

$q(y)$ の台は有界ではない

$$R(D) = \frac{1}{\beta} \log \frac{1}{\alpha D} \quad (0 \leq D \leq 1/\alpha)$$

- $\beta < \beta' < 2$ では, $q(y)$ は離散分布 (かつ有限)
[Dytso et al., 2018]



一般情報源, 二乗損失 ($R(D) > R_L(D)$)

- 最適再構成分布は離散分布 [Fix, 1978] [Rose, 1994]

Y の台が区間 I_0 を含むとする。(「集積点を持つとする。」に拡張可)

$$\rightarrow \forall y_0 \in I_0 \quad c(y) = 1 \iff \int \frac{p(x) \exp(s(x - y_0)^2)}{\int q(\tilde{y}) \exp(s(x - \tilde{y})^2) d\tilde{y}} dx = 1$$

$$\rightarrow \int \frac{p(x) \frac{\partial^n}{\partial y_0^n} \exp(s(x - y_0)^2)}{\int q(\tilde{y}) \exp(s(x - \tilde{y})^2) d\tilde{y}} dx = 0 \quad \begin{array}{l} \text{ここで} \\ \frac{d^n}{dz^n} e^{sz^2} = H_n(z) e^{sz^2} \end{array} \quad \begin{array}{l} \text{エルミート多項式} \\ \downarrow \end{array}$$

n回微分

$$\iff \frac{p(x)}{\int q(\tilde{y}) \exp(s(x - \tilde{y})^2) d\tilde{y}} \quad \text{が全ての(一次以上の)エルミート多項式と直交, すなわち定数}$$

$$\iff p(x) = \sqrt{\frac{|s|}{\pi}} \int q(\tilde{y}) \exp(s(x - \tilde{y})^2) d\tilde{y} \iff R(D_s) = R_L(D_s)$$

ガウス情報源, 絶対損失 [Tan, Yao, 1975]

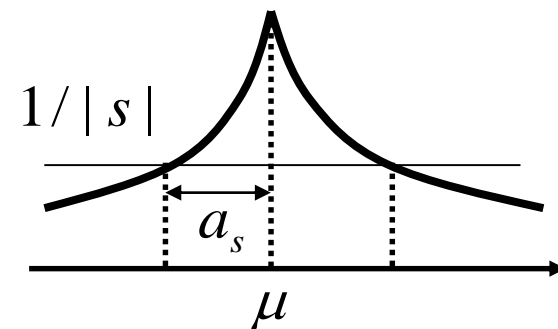
○ $p(x) = N(x | \mu, \sigma^2)$ $d(x, y) = |x - y|$

➡ $R(D) > R_L(D)$

$V_s = [\mu - a_s, \mu + a_s]$ と仮定

$c(y) = 1$ ($y \in V_s$) (q(y)について解く)

↔ $\frac{1}{|s|} = \frac{\text{erfc}(a_s / \sigma)}{p(\mu - a_s)} = \frac{\text{erfc}(a_s / \sigma)}{p(\mu + a_s)}$



$q_s(y) = \left\{ 1 + \frac{1}{s^2 \sigma^2} - \frac{(y - \mu)^2}{s^2 \sigma^4} \right\} p(y)$ 連続成分

$+ \frac{1}{s^2} \left(|s| - \frac{a_s}{\sigma^2} \right) p(y - a_s) \{ \delta(y - \mu + a_s) + \delta(y - \mu - a_s) \}$ 離散成分

$c(y) \leq 1$ ($y \notin V_s$) ➡ $R(D)$ (sによるパラメータ表示)

ガンマ情報源, 対数絶対損失 [Watanabe, Ikeda, 2016]

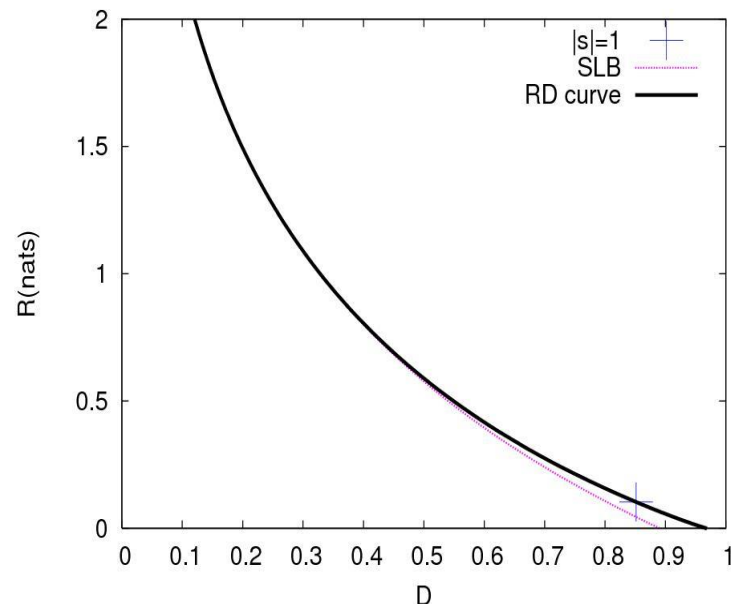
- ガンマ情報源

$$p(x) \propto x^{\alpha-1} e^{-x}$$

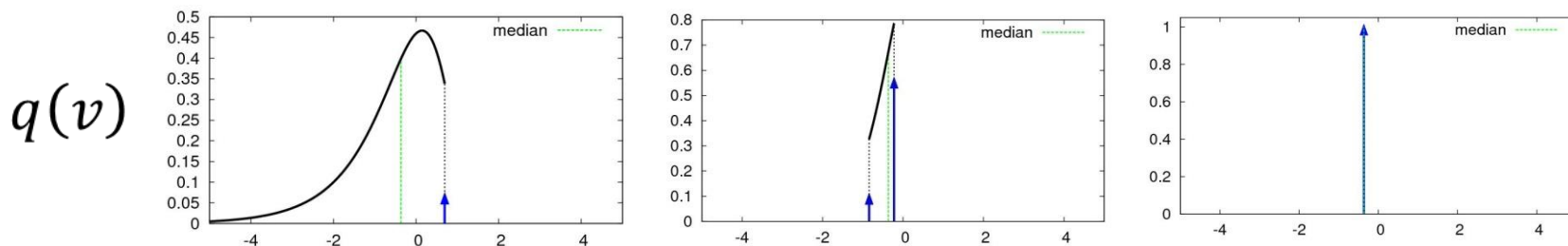
- 音声信号処理, トラフィック理論 ...

- 対数絶対損失

$$d(x, y) = \underbrace{|\log x|}_u - \underbrace{|\log y|}_v$$



□ 最適再構成分布 (連続 + 離散)

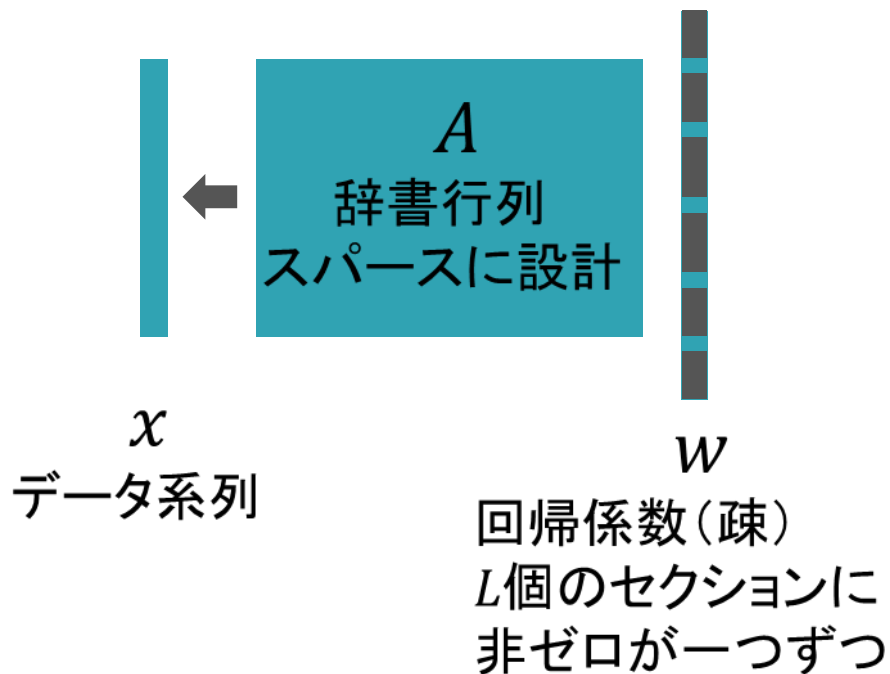


Dが小: 左側の離散成分が消える

Dが大

スパース回帰符号

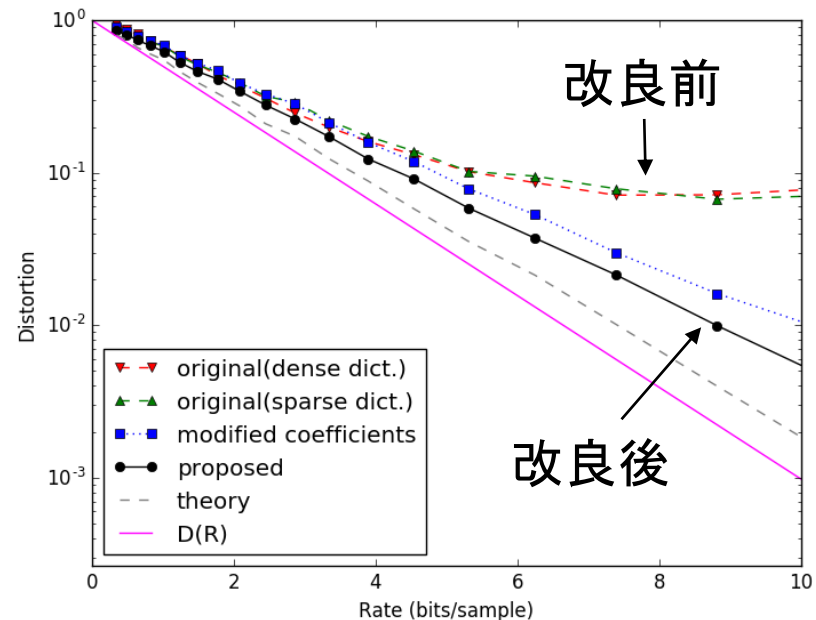
- スパース回帰符号化法:
 - スパース推定に基づく有歪圧縮法
 - 二乗損失において $R(D)$ を達成 [Venkataramanan et al., 2014]



損失関数の拡張: 絶対損失



□ 最適再構成分布に基づいた辞書行列, 回帰係数の設計 [Konabe, Watanabe, 2018]



まとめ

- レート歪み関数と最適再構成分布の解析
 - べき乗損失のレート歪み関数
 - 対数絶対損失のレート歪み関数 ($R(D) > R_L(D)$)
 - 絶対損失下でのスパース回帰符号