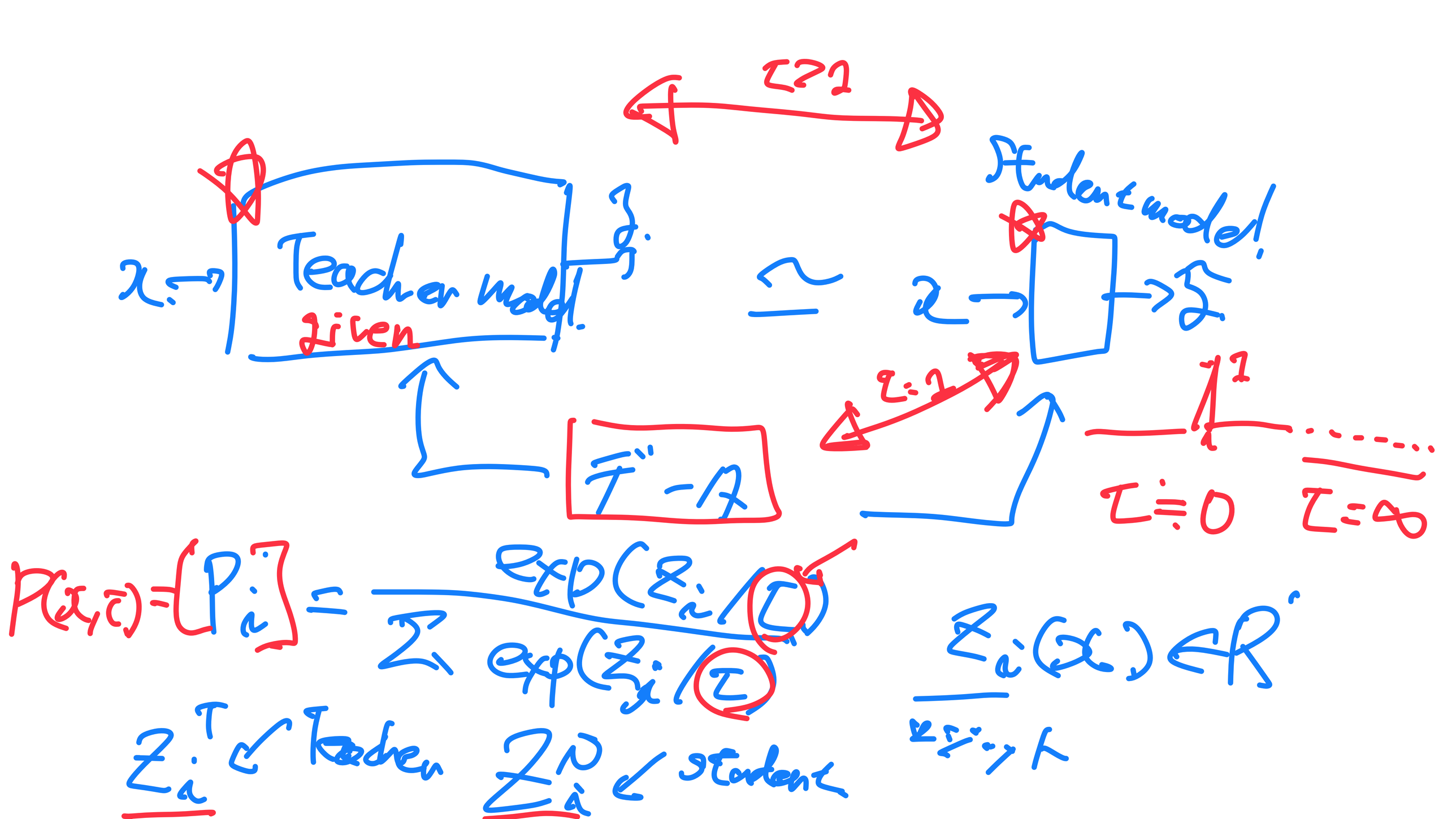
2025.8.18

転移学習

知識蒸留、マルチタスク

京大 岡本有司


τ>1
Teacher model
given
x
y
≃
Student model
z
ζ
τ=1
τ=0
τ=∞
T-A
$P(x,\tau)=[P_i]=\frac{\exp(z_i/\tau)}{\sum_j \exp(z_j/\tau)}$
$z_i(x)\in R^n$
ロジット
z_i^T teacher
z_i^S student

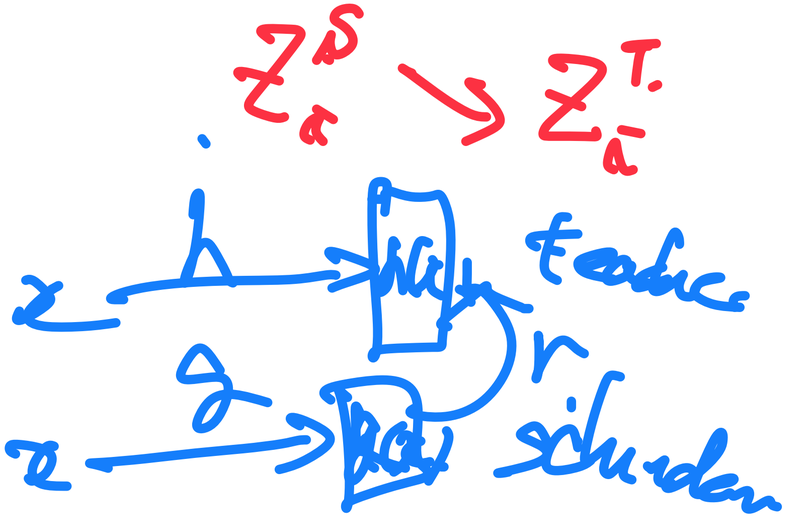
$$Loss = \alpha L_D(p(z_T, \tau), p(z_S, \tau)) + (1-\alpha) L_S(y, p(z_S, \tau))$$

$$L_S = \sum_{i=1}^{C} -y_i \log p(z_{S_i}, \tau)$$

$$L_D = \sum_{i=t}^{C} -p(z_{t_i}, \tau) \log p(z_{S_i}, \tau)$$

$$L_{DR} = \tau^2 KL(p(z_S) | p(z_T))$$

$$L_{DF} = \| f^s(x) - r \circ g(x) \|^2$$



蒸留の分類

Loss 関数のちがい

§7.2

1. 応答ベース

2. 特徴ベース
（中間出力）

3. 関係ベース
（どうけい関係）

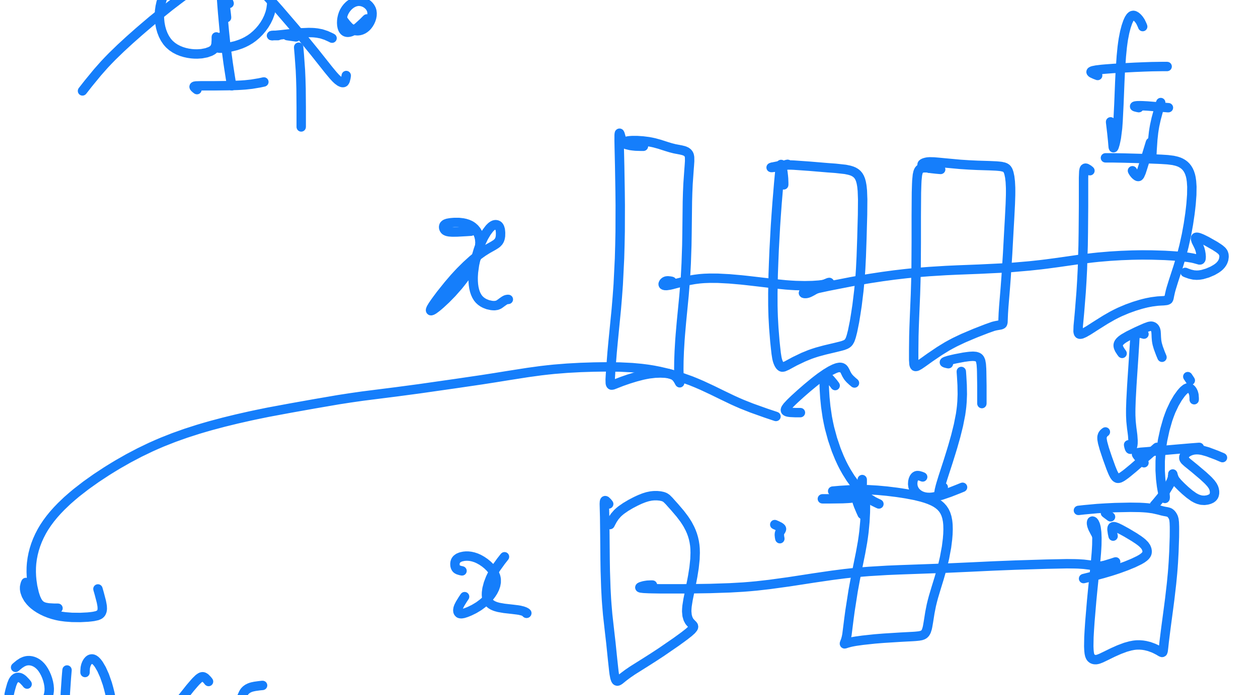
学習アルゴリズムがちがう

1. オフライン　teacher fix

2. オンライン → (teacher 1 ⇕ student ‖ teacher2)

3. 自己蒸留 → One-teacher モデル

$$L_{DF} = \ell\left(f_T(x), \Phi_S \circ t_{S(x)}\right)$$



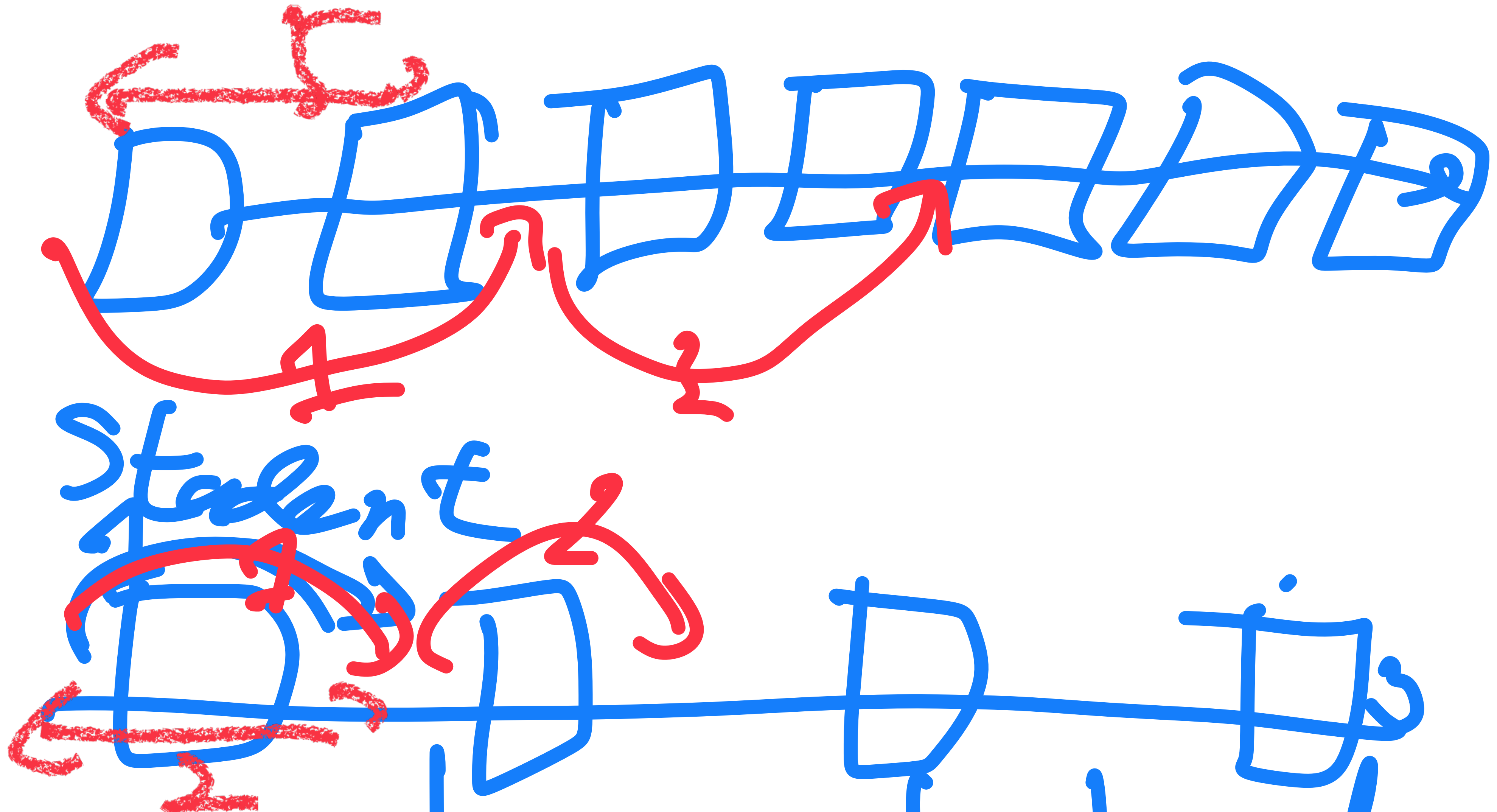
3. 関係

$$\mathcal{L}_{rel} \triangleq L\left(\psi_T(\hat{f}_T, \check{f}_T), \psi_S(\hat{f}_S, \check{f}_S)\right)$$

Attention map

teacher

student



FSP

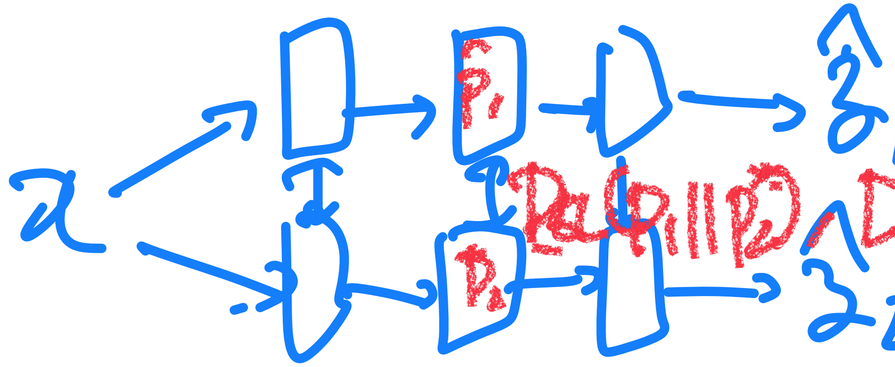
$$G_{ij}^{S} \triangleq \sum_{k=1}^{K}\sum_{l=1}^{L} \frac{f_{kli}^{S}(x)\, \tilde{f}_{klj}(x)}{KL}$$

$f^S \in \mathbb{R}^{K \times L \times n}$: K : width, L : height, n : feature dim.

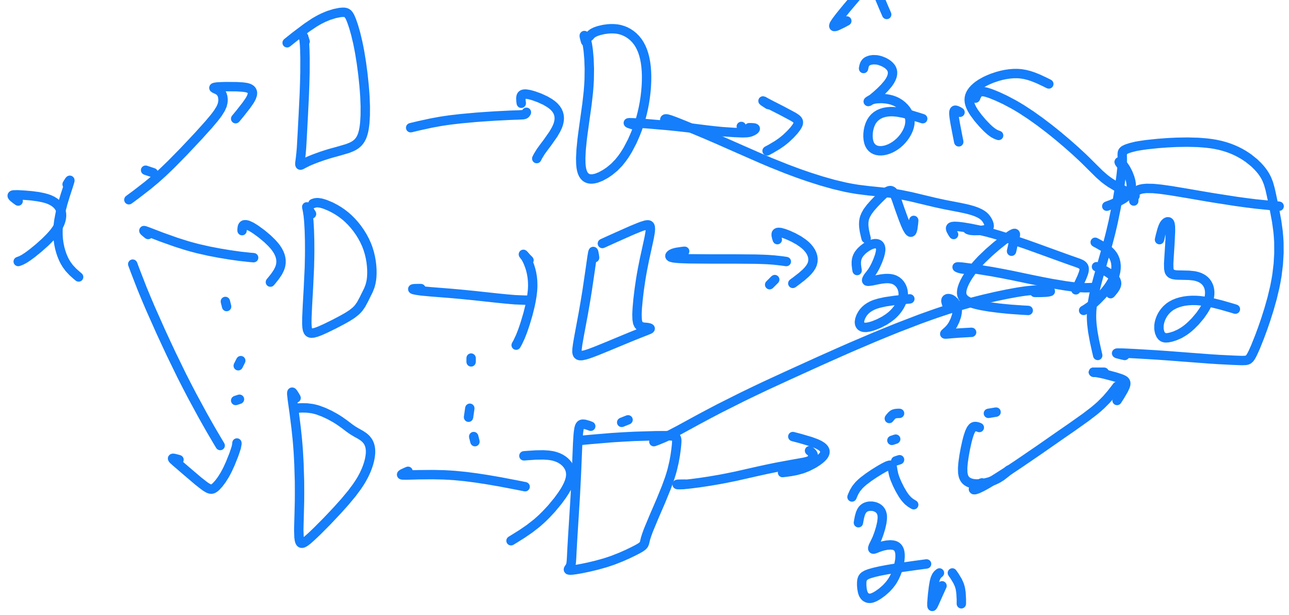
$$\mathcal{L}_{SFP} \triangleq \frac{1}{n}\sum_{i=1}\sum_{j=1} \lambda_j \left\| G^T(x_i) - G^S(x_i) \right\|_2^2$$

オンライン型

- DML
- ONE



※モデルのバラツキを
おさえる.

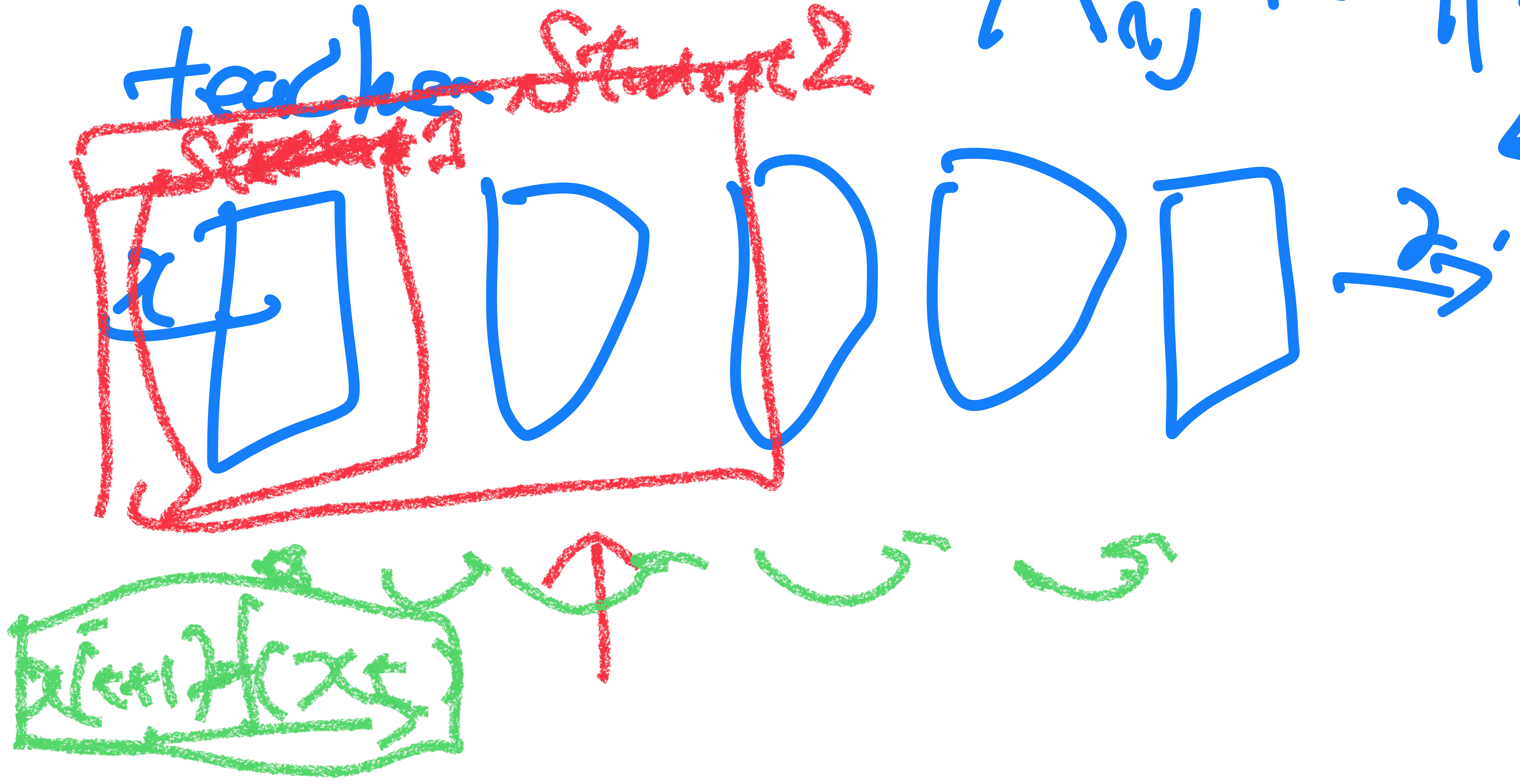


自己蒸留

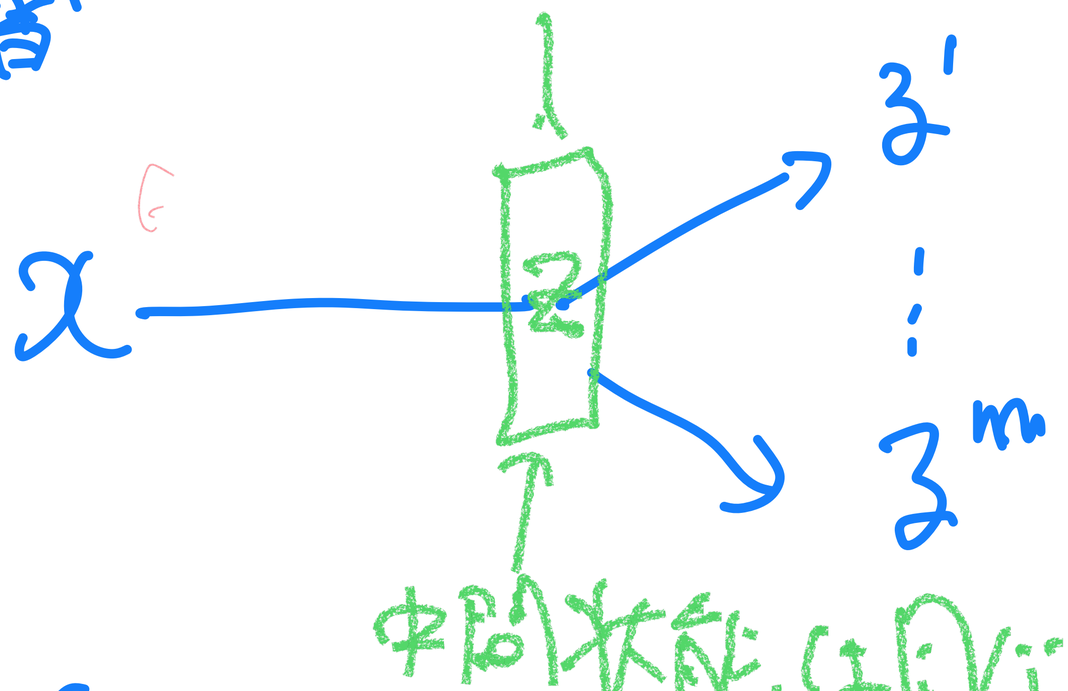
統系　　$s' = s+1$　　☆ モデルのバラツキをおさえる.

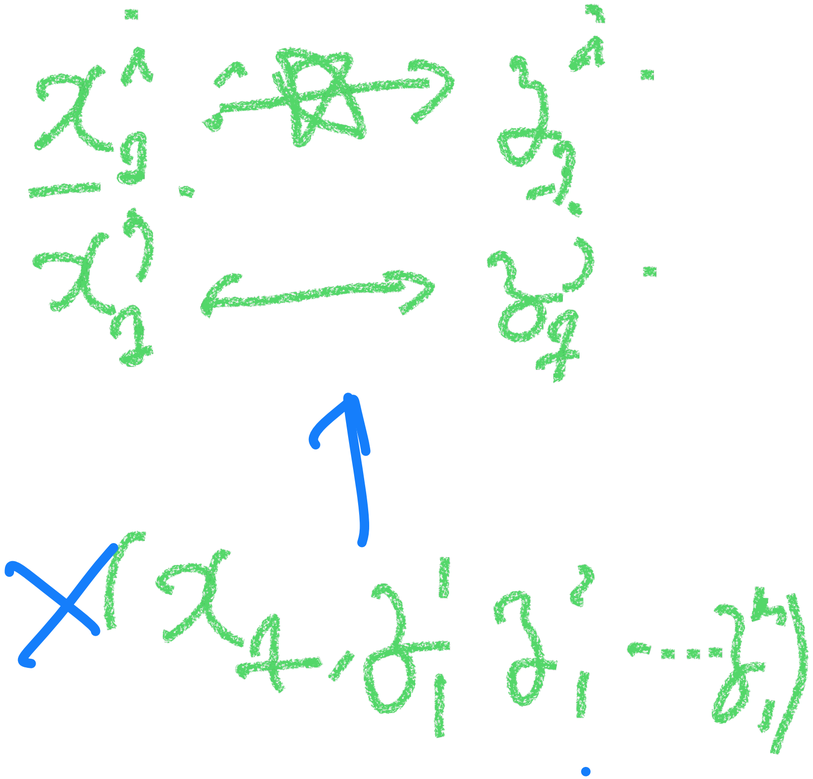
$$\mathcal{L}_D \triangleq l\left(\Psi(A^{s}) - \Psi(A^{s'})\right)$$

$$A^{s}_{ij} = \| f^{s}(x_i) - f^{s}(x_j) \|_2$$



マルチタスク学習





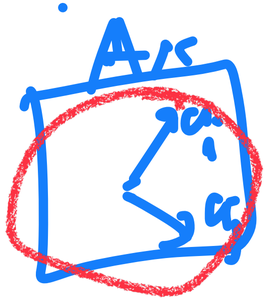
特徴変換 (z=x)

$$\min_{W,b} L(W,b) + \lambda \|A\|_{2,1} \quad \text{s.t. } U^T U = I$$

$$L(W,b) = \sum_{i=1}^{m} \sum_{j=1}^{n} l(y_j^i, W x_j^i + b)$$

$z_j^i = W x_j^i + b$

$u_i = Ax$



§8.2.2 タスク クラスタリング

$$\min_{S,L,b} \mathcal{L}(L,S,b) + \lambda_1 \|S\|_1 + \lambda_2 \|L\|_F^2 \quad \text{s.t.} \quad U^T U = I$$

$$\mathcal{L}(L,S,b) \triangleq \mathbb{E}\left[\|y_i - L_i S f_\theta(x_i)\|^2\right]$$ ← タスク

§8.2.3 パラメータ分解

$$\min L(W,b) + \sum_{k=1}^{h} g_k(W_k)$$

$W = \sum_{k=1}^{h} W_k$ ← v.s. たし算

ex.

$g_1(W) = \|W\|_{2,1}$

$g_2(W) = \|W\|_2$