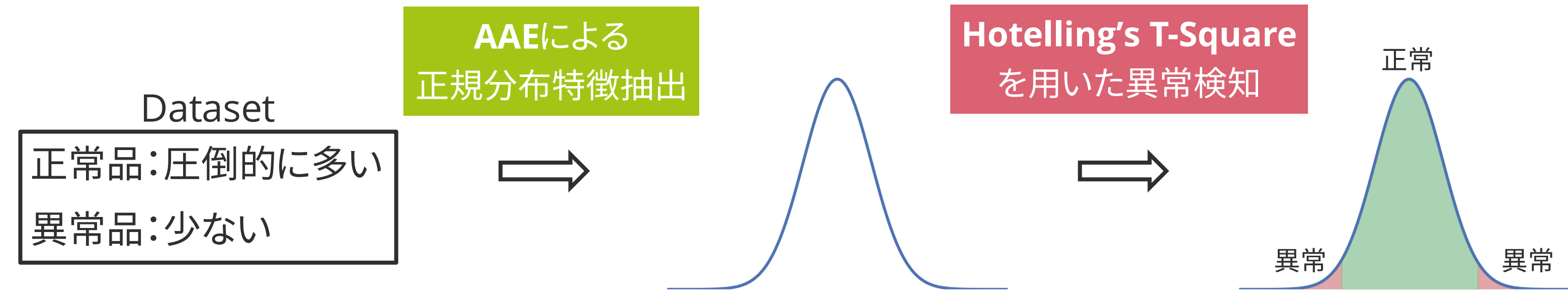


少数不良品サンプル下における深層学習による正常モデル生成と異常品検出

工学部 電気電子・情報工学科 情報コース 加藤 邦人
kkato@gifu-u.ac.jp

アイデア

- 不良品が少ない状況において、有効なDeep Learningを用いた外観検査手法
- AAE(Adversarial AutoEncoder)による正規分布に従う低次元特徴量の抽出
- 抽出された正規分布特徴から異常度を算出、異常度に対して1次元のしきい値設定



Hotelling's T-Square

- 正規分布に従っているデータに対して強力な異常検知手法
- 平均 μ と分散 Σ から未知データの異常度(マハラノビス距離)を推定
- 異常度は、サンプル数が特徴次元数より圧倒的に大なら、カイ2乗分布に従う

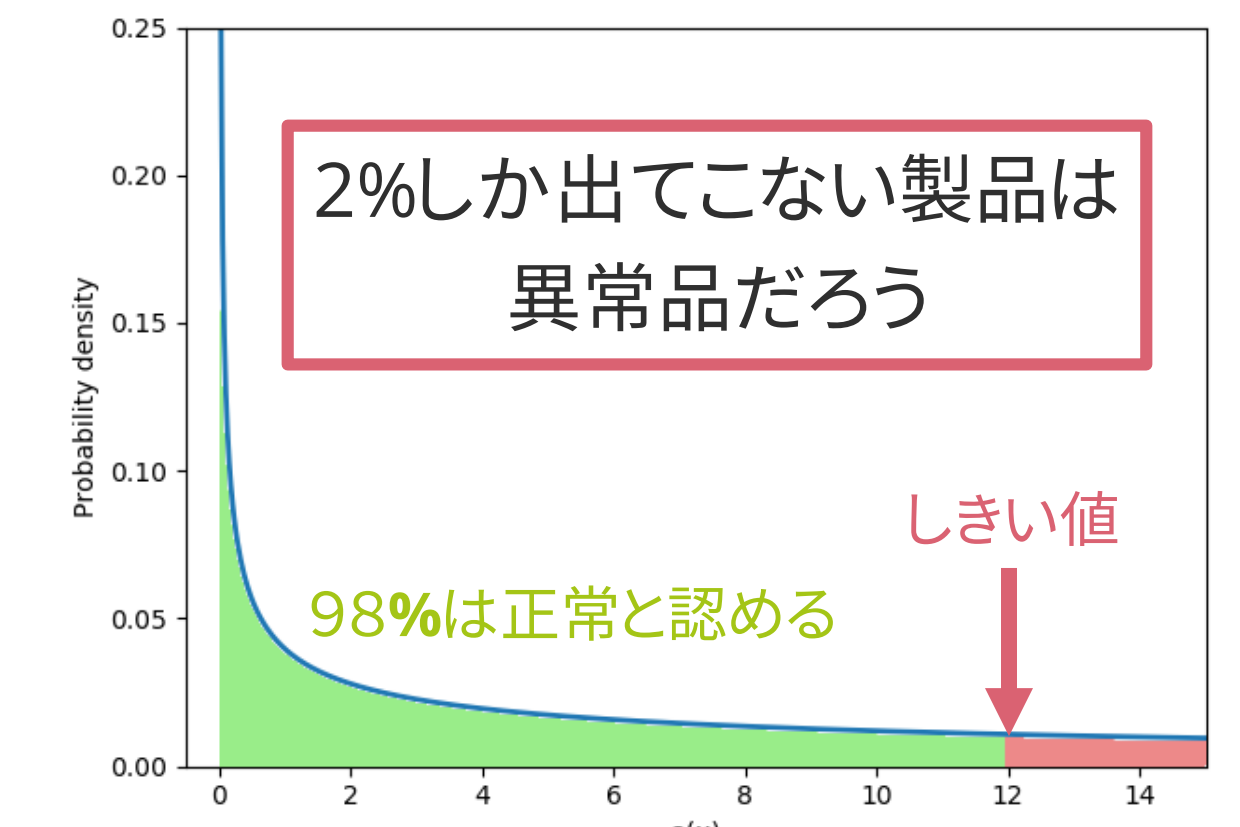
未知データ x' に対する異常度

$$a(x') = -\ln N(x' | \mu, \Sigma)$$

$$\propto (x' - \mu)^T \Sigma^{-1} (x' - \mu)$$

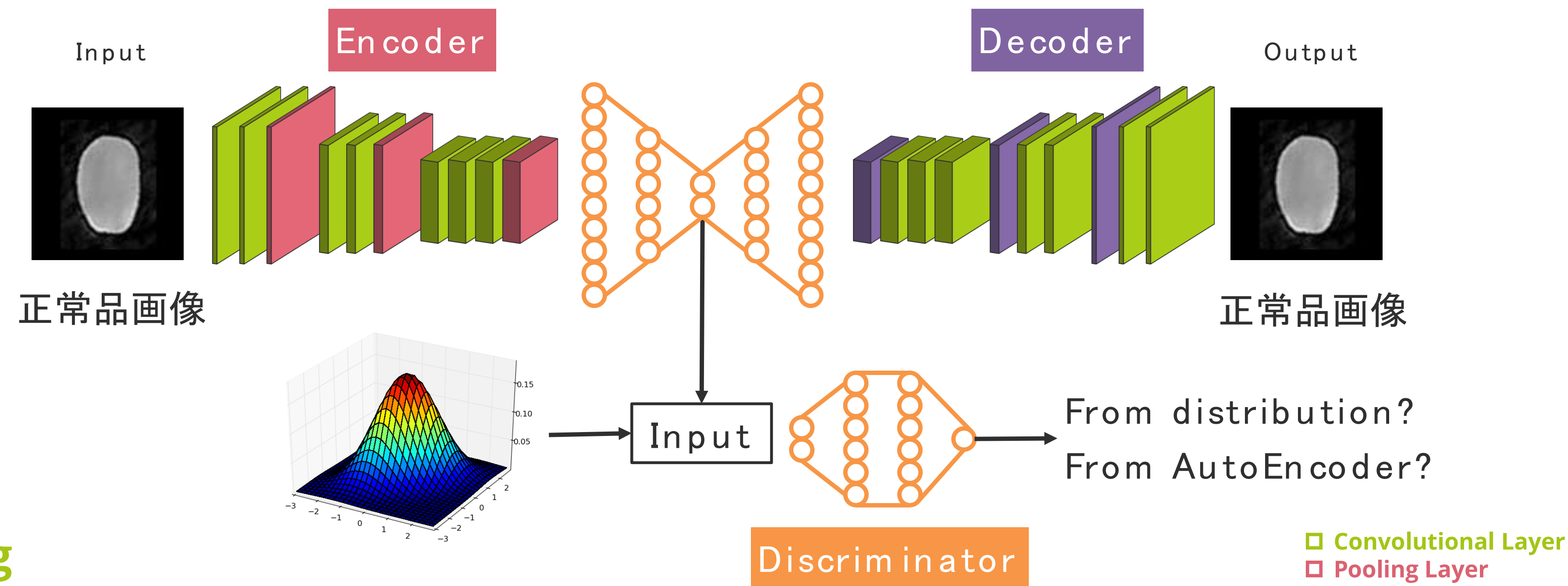
マハラノビス距離

x	: training data
μ	: mean vector
Σ	: covariance matrix
x'	: unknown data



Adversarial AutoEncoder

- 特徴抽出に用いられるAutoEncoderがベースモデル
- AutoEncoderに敵対的学習を組み込んだ枠組み
- 抽出する特徴量を任意の分布に従わせることができる



Training

- Reconstruction phase
- 入力をうまく再構成できるようにMSEを最小化するようにAutoEncoderを更新

$$L_{AE} = \frac{1}{2} \sum_{i=1}^m (x_i - De(En(x_i)))^2$$

- Regularization phase
- 対象の分布由来なのか, AutoEncoder由来なのかを

Discriminatorがうまく識別できるようにDiscriminatorを更新

$$L_{Dis} = -\left[\frac{1}{m} \sum_{i=1}^m \log D(z_i) + \frac{1}{m} \sum_{i=1}^m \log(1 - D(En(x_i))) \right]$$

- Discriminatorを騙すように、抽出する特徴を
- 対象の分布からサンプリングされたようなベクトルに近づけるようにEncoderを更新

$$L_{En} = -\left[\frac{1}{m} \sum_{i=1}^m \log(D(En(x_i))) \right]$$

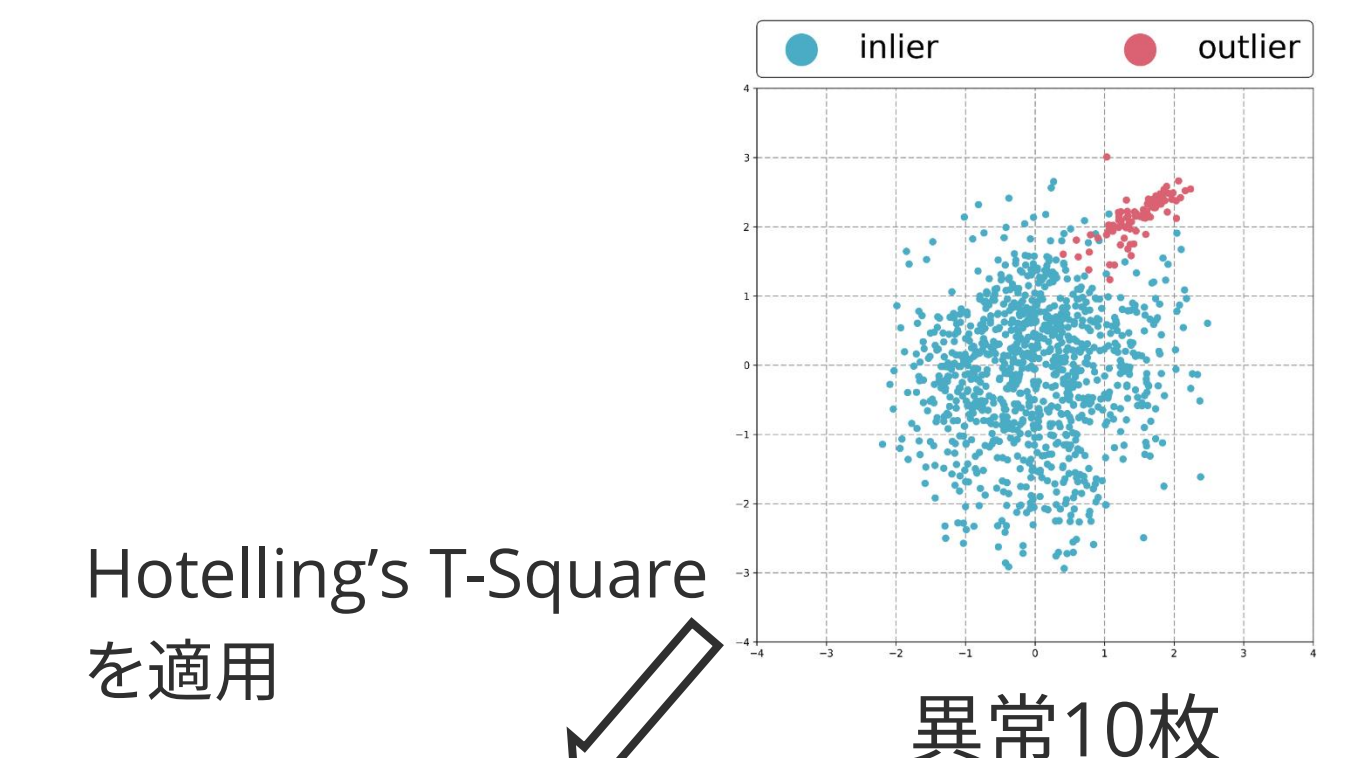
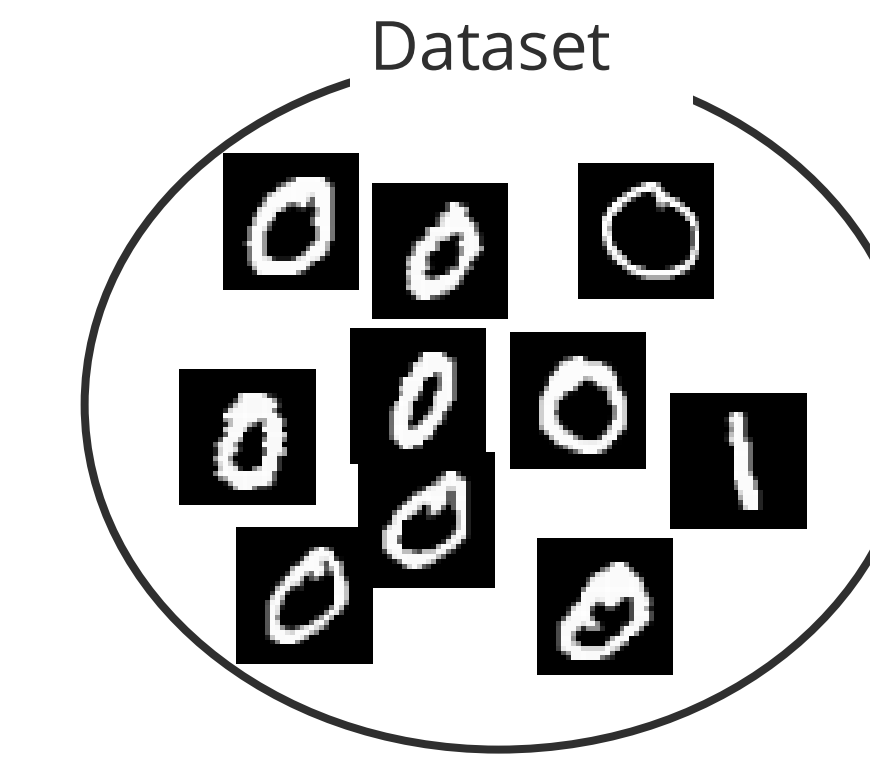
x : input image
 $De(\cdot)$: Decoder
 $En(\cdot)$: Encoder
 m : Minibatch size
 z : noise vector
 $D(\cdot)$: Discriminator

Convolutional Layer
 Pooling Layer
 Upsampling Layer
 Reshape & FC

実験

MNIST

- 正常データ"0" (4950枚)
- 異常データ"1" (10, 100枚)
- 2次元の標準正規分布



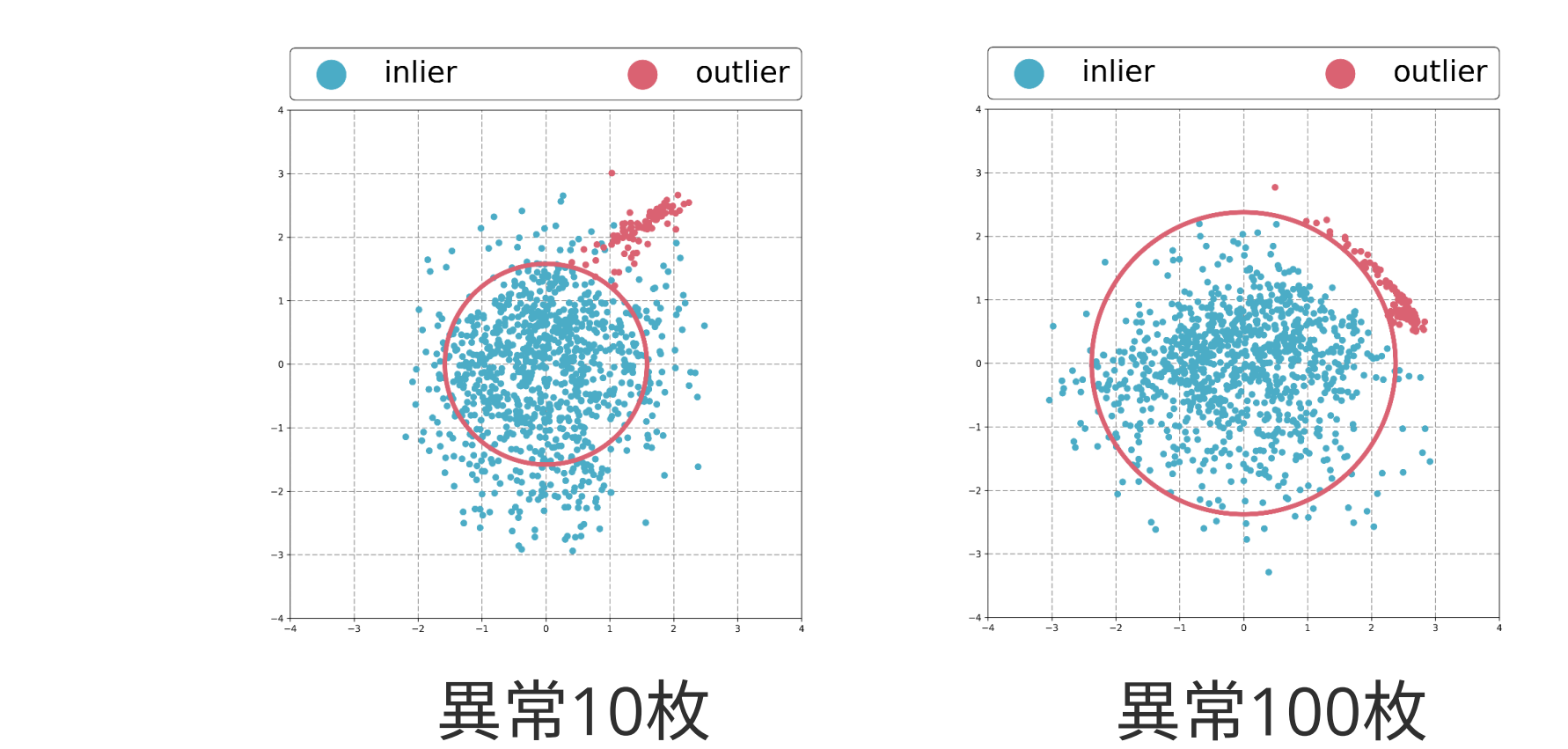
標準正規分布の際の異常度

$$\mu : \mathbf{0}$$

$$\Sigma : I$$

$$a(x') \propto x'^T x'$$

原点からのユークリッド距離



赤円は、異常の見逃しが0となるようなしきい値

白米データセット

- 29164枚の白米データセット
- 6%ほどの異常が含まれている
- 64×64のグレースケール画像
- 16次元の標準正規分布

正常データ



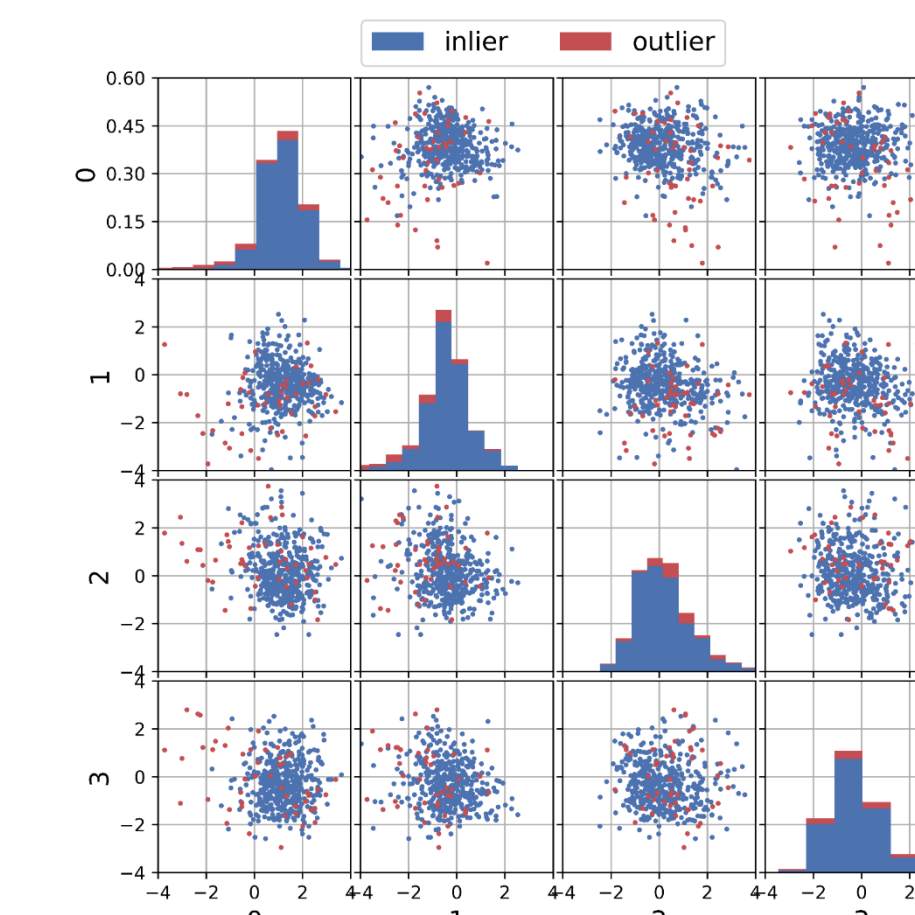
異常データ



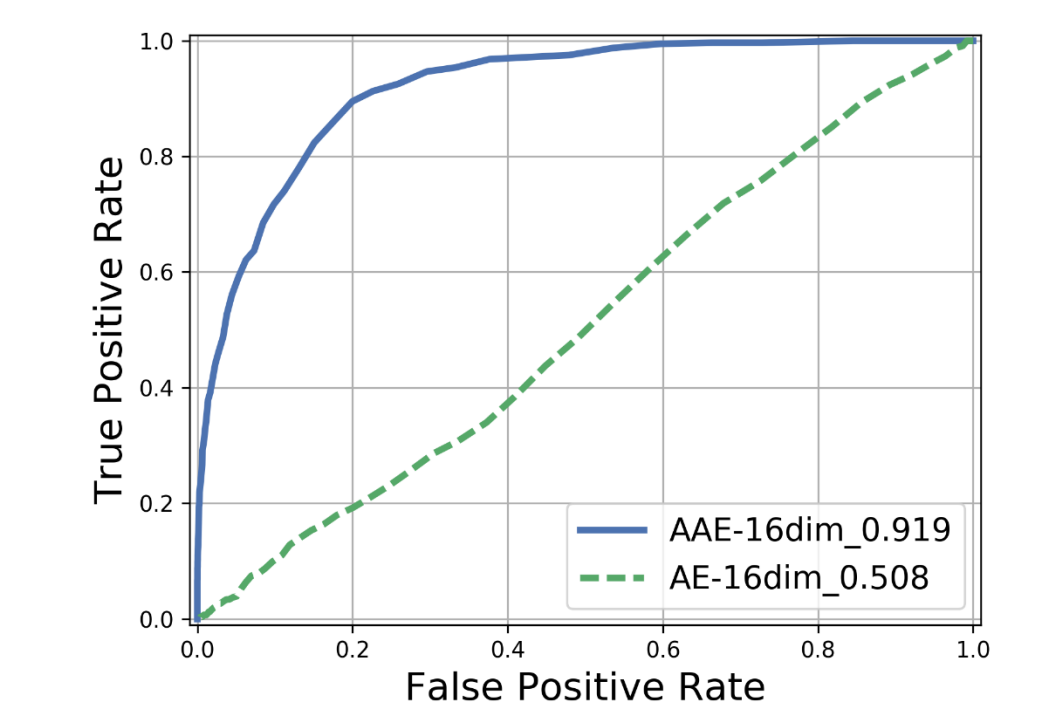
異常度が小さいデータ



異常度が大きいデータ



抽出された
1~4次元目の特徴



ROC曲線
AUC: 91.9%