Beyond cross-entropy: learning highly separable feature distributions for robust and accurate classification

Arslan Ali, Andrea Migliorati, Tiziano Bianchi, Enrico Magli
Politecnico di Torino

GCCS Architecture

- GCCS: Gaussian class-conditional simplex loss
- High Separability between classes
- Extract discriminative features from the input data
- Map the features to well behaved, target Gaussian distributions

GCCS loss

 Authorized users loss

\[ \mathcal{L}_i = \frac{1}{i} \left[ \log \left( \frac{\Sigma_{T_i}}{\Sigma_{O_i}} \right) - D + \text{tr} \left( \Sigma_{T_i}^{-1} \Sigma_{O_i} \right) + (\mu_{T_i} - \mu_{O_i})^T \Sigma_{T_i}^{-1} (\mu_{T_i} - \mu_{O_i}) \right] \]

\[ \mathcal{K}_i = \left( \frac{x - \mu_{O_i}}{\sigma_{O_i}} \right)^4 \]

Total loss: \[ \mathcal{L}_{\text{GCCS}} = \sum_{i=1}^{D} \left[ \mathcal{L}_i + \lambda (\mathcal{K}_i - 3) \right] \]

GCCS decision rule

- Partition the decision space into Voronoi regions
- Compute the distance from all the centers - choose minimum

\[ \hat{y} = \arg \max_i z_i, \]

- Index of the predicted class for the test image

Fig. 1: The GCCS architecture takes input data and learns discriminative features that are mapped onto Gaussian target distributions in the latent space.