DL L17: 07.07.2022

(-) need paired data $(\underline{x}, \underline{y})$, costly or sometimes impossible

(c) Solution: Cycle GAN with unpaired data

9-30
- no L1 loss $\|\underline{x} - \hat{\underline{x}}\|_1$, $\|\underline{y} - \hat{\underline{y}}\|_1$ → relaxes data collection

9-31

10 New trends and old fashions

10.1 ~ 10.2 : New trends
10.3 ~ 10.6 : "old" relevant NNs and modules

10.1 Self-Supervised Learning (SSL)

Learning with supervised methods but unlabelled data

10-1

10-2

Different concepts of SSL

10.1.1 Pretext task

- dont learn the final ~~tax~~ ML model; No labels, no task
- learn a representation / features of input which are useful for future downstream tasks
- by solving pretext tasks

10-3

10-4

10-5

Pre-training

↳ all the network + weights stored in Tensorflow as they are already trained on various places

Key step:

Design of the PT

E 10.1

Auto-encoder

PT: Self-reconstruction : $\tilde{x} = y = x$

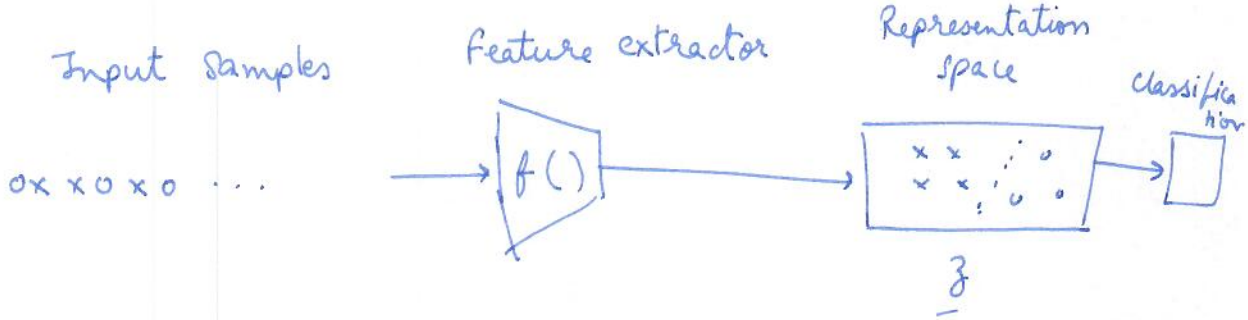
10-6

10-7

10.1.2 Contrastive Learning

Learn to contrast

Idea



Goal:

Learn the feature exactor $f()$ such that

- similar samples (+ves) stay close to each other } contrast
  dissimilar " (-ves) " far from " "

→ Simpler final classification

We generate similar / dissimilar samples from unlabeled data

10 - 8

10 - 9 shows the data augmentation

Steps

- minibatch : B, randomly selected images $\underline{x}(1) \cdots \underline{x}(B)$ unlabeled

- pairwise data augmentations to get similar pairs with similar content but different data, NOT to remove overfitting

  ⟶ 2B views

  $\tilde{\underline{x}}_{2i-1} = t(\underline{x}_i)$, $\hat{\underline{x}}_{2i} = t'(\underline{x}_i)$, $t = t'$, $1 \le i \le B$

  $(\tilde{\underline{x}}_{2i-1}, \hat{\underline{x}}_{2i})$ = positives

  remaining 2B - 2 views : negatives to $(\tilde{\underline{x}}_{2i-1}, \tilde{\underline{x}}_{2i})$

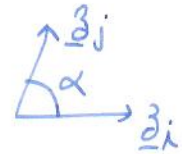
- representations : $\underline{h}_i = f(\tilde{\underline{x}}_i)$ $1 \le i \le 2B$

- projections : $\underline{z}_i = g(\underline{h}_i)$ $1 \le i \le 2B$

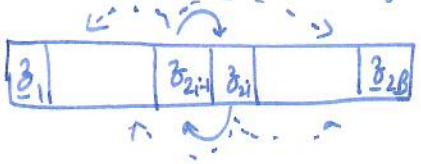
- similarities $a_{ij} = s(\underline{z}_i, \underline{z}_j)$ $1 \le i \ne j \le 2B$

  eg cosine similarity

  $$s(\underline{z}_i, \underline{z}_j) = \frac{\underline{z}_i^T \underline{z}_j}{\|\underline{z}_i\| \, \|\underline{z}_j\|} = \cos\alpha$$



- Softmax likelihood for positives



→ +ve → so high value

⋯→ -ve → low value

this is the top pass

$$q_{2i-1,2i} = \frac{e^{a_{2i-1,2i}}}{\sum_{k=1}^{2B} e^{a_{2i-1,k}}}$$

$k \ne 2i-1$

$$q_{2i,\,2i-1} = \frac{e^{a_{2i,\,2i-1}}}{\sum_{\substack{k=1 \\ k \neq 2i}}^{B} e^{a_{2i,k}}} \in [0,1]$$

$$\neq q_{2i-1,\,2i}$$

- negative log-likelihood (NNL) see ch 3.4

$$l_{2i-1,\,2i} = -\ln q_{2i-1,\,2i}$$

$$l_{2i,\,2i-1} = -\ln q_{2i,\,2i-1}$$

- minibatch loss

$$\mathcal{L} = \frac{1}{2B} \sum_{i=1}^{B} \left(l_{2i-1,\,2i} + l_{2i,\,2i-1} \right)$$

$$\min_{f(),\,g()} \mathcal{L}$$

Q Won't the f.n.s $f()$ and $g()$ contrast each other? i.e. if we try to minimize $\mathcal{L}$

10.2 Attention and transformer

10.2.1 Attention

Attention : Focus on relevant regions of data

$\hat{=}$ correlations in signal processing

10 - 12

"Everything is attention" → attention and transformer course of Uni Stanford.