

TTIC 31230, Fundamentals of Deep Learning

David McAllester, Winter 2020

Contrastive Predictive Coding

Contrastive Predictive Coding (CPC)

We consider a population distribution on pairs $\langle x, y \rangle$.

For example x and y might be video frames separated by 10 seconds in a video.

For simplicity we will assume that the marginal distributions on x and y are the same — the probability that an image occurs as a first frame is the same as the probability that image occurs as a second frame.

In CPC we draw a pair $\langle x, y \rangle$ and **minimize** a discriminator loss for distinguishing $z_\Phi(y)$ from $z_\Phi(\tilde{y})$ for $\tilde{y} \sim \text{Pop}(y)$. The discriminator gets to see x .

Contrastive Predictive Coding (CPC)

For $N \geq 2$ let $\tilde{P}^{(N)}$ be the distribution on tuples $\langle i, y_1, \dots, y_N, x \rangle$ defined by the following process.

- draw a pair $\langle x, y \rangle$ from the population.
- drawn a sequence of $N - 1$ “distractor values” from the marginal distribution $\text{Pop}(y)$. These are unrelated to x .
- insert y at a random position among the distractors to get the sequence y_1, \dots, y_N .
- return the tuple $\langle i, y_1, \dots, y_N, x \rangle$ where i is the index of y among the distractors.

Contrastive Predictive Coding (CPC)

$$\Phi^* = \underset{\Phi}{\operatorname{argmin}} \mathcal{L}_{\text{CPC}}(\Phi)$$

$$\begin{aligned} \mathcal{L}_{\text{CPC}}(\Phi) = E_{\langle i, y_1, \dots, y_N, x \rangle \sim \tilde{P}^{(N)}} \\ - \ln P_{\text{CPC}}(i | z_{\Phi}(y_1), \dots, z_{\Phi}(y_N), z_{\Phi}(x)) \end{aligned}$$

$$P_{\text{CPC}}(i | z_1, \dots, z_N, z_x) = \underset{i}{\operatorname{softmax}} \ z_i^{\top} z_x$$

Contrastive Predictive Coding (CPC)

$$\Phi^* = \underset{\Phi}{\operatorname{argmin}} \mathcal{L}_{\text{CPC}}(\Phi)$$

$$P_{\Phi}(i|z_1, \dots, z_n, z_x) = \underset{i}{\operatorname{softmax}} z_i^{\top} z_x$$

As N gets larger the contrastive discrimination task gets harder.

The task is also made difficult by the requirement that the score is defined to be an inner product of feature vectors.

Contrastive Predictive Coding (CPC)

(SimCLR:) A Simple Framework for Contrastive Learning of Visual Representations, Chen et al., Feb. 2020 (self-supervised leader as of February, 2020).

They use a distribution on pairs $\langle x, y \rangle$ defined by drawing an image s from ImageNet and then drawing x and y as random “augmentations” (modifications) of the image s — either a random translation, rotation, color jitter, masking, edge image, or a composition of these modifications.

Contrastive Predictive Coding (CPC)

The feature map z_Φ can then be applied to the images of ImageNet.

The feature map z_Φ is then tested by using a **linear** classifier for ImageNet based on these features.

SimCLR

END