IDL - Appeal, Exam B.

David Ponarovsky

August 2025

Question 7 - RNN nets. Does a recursive net of type Elman, that gets the zero vector as input at each step, can count? Namely, outputs the value t at its t-th step?

1. Yes
2. No
3. In general No, Yet when given t as initial input, yes.

My answer: (3), Correct answer: (1) .

I believe that the confusion arises from the order of entities. The question: 'Is there an Elman cell that can count until t for an arbitrary t?' is a different question than: 'Fix t, does there exist an Elman cell that can count until t?'

For the **second** question: There is a family of unbounded fan-in/out circuits, at width $poly(|t|)$ (the length of the encoding of t), that implement addition: [addition in AC_0]. It's not hard to see that the implementation in the notes can be realized using an Elman cell, and therefore one can find such a realization that adds 1 to the input which is entered via the hidden channel.

For the **first** question, in which we first fix the Elman cell architecture, and therefore also fix the width of the output. Thus, the number that can be outputted using the cell has to have a suitable length and cannot be arbitrary.

I marked the third option since saying 'on given t as initial input' implies that t is a valid input/token to the cell, namely, it has a length that matches the cell's weights. For such t's, the RNN counts correctly, whereas for others, the behavior of giving the input is not well defined and is expected to be a failure.

Question 14 - VAEs. What is the reason for the generated images by VAEs been blurred compared to the images generated by GANs ?

1. Usage of reconstruction loss that smooth sharp items.
2. KL-divergence element that impair the disentangle (or separation) of different samples in the latent space.
3. Low presentation ability of the VAEs architecture.

4. Entering too mach noise into the latent space, which after decoding comes into fact in burred image.

My answer: (1), Correct answer: (2) . I agree that, in general, the main reason for the results of the VEAs being blurred is the KL-divergence term. In particular, it enforces the decoder to decode a sampled superposition over the latent space. Yet the question asks what is the main reason for blurriness compared to GANs.

When comparing to GANs, which, in our course, have an amorphic architecture and can be arbitrarily complex, one should also consider the case when the architecture of the VEAs is exactly as complicated as in GANs, surely if in that regime the VEAs products are still inferior.

In that regime, one could think of a **decoder** which expands the latent dimension so much, such that for the human eye (or more correctly to the latent-space eye) the interpolation $D(tz_1 + (1-t)z_2)$ is not a continuous function. In that case, even though the KL-element enforces the latent vector z to be distributed according to a Gaussian, any value of z that can be seen in the experiment leads to another (isolated) value x in the data space.

For example, consider that the latent space has a width of k bits (for the sake of the exercise, you can imagine that the entities in the latent space are numbers in $(0,1)$), and the decoder decodes each z to a space represented by $n = 2^k$ bits, such that $D(z) = 2^{z*2^k}$ (shifting by the integer suitable to z). Clearly, close samples in the latent space get far in the domain space. That argument cancels any justification due to behavior or penalization in the latent space.

That brings us to ask if there is a difference in the case where the latent space is trivial. Namely, when its dimension is 0 (its size is 1), namely a constant machine ('generates only a cat, and always the same cat'). In that case, the optimal generator and discriminator in GANs would be the generator which outputs the same cat, and the discriminator which guesses at probability $\frac{1}{2}$, while in the VEA scheme any generator which outputs a noisy version of the same cat would be penalized by only $\frac{\varepsilon}{\sqrt{n}}$ if the reconstruction error is l_2 and noise is distributed independently over the pixels with mean ε. That loss is going to zero when the domain dimension is increasing. (If that reconstruction loss is an l_1 loss, then one should use a correlative noise mode, for example, one that hits $\sqrt{n}$ of the bits on average, which matches the blurring of the perimeter of the cat.)

To complete that argument, we have to show in the GANs setting that generating noisy versions of cats would have a non-trivial loss cost. We show that by defining the following simple discriminator. On a given $\tilde{x}$, take the subtraction: $\Delta \leftarrow \tilde{x} - \text{The-Cat}$. If Δ is non-zero, then mark $\tilde{x}$ as a generated image.

Hence, we showed that the source for GANs to output sharper images than their complex-equivalence VAEs emanates from the choice of the reconstruction loss.