Deep Learning & Applied AI

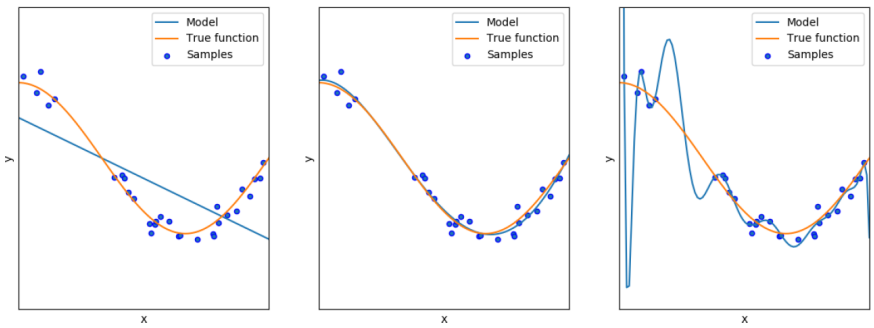
April 26, 2023

Your name:

Your matr. number:

Question 1 (3 points) Prove or disprove: deep ReLU networks are linear.

Question 2 (5 points) Consider the following 2D plots:



These capture three different phenomena – which ones? Please explain to the best of your understanding.

Question 3 (8 points) Consider a deterministic (i.e. not variational) autoencoder. How would you modify it so that it behaves like PCA? You are free to modify the architecture, the loss, or both.

Question 4 (5 points) The encoder-decoder model of an AE is strictly related to the notion of *manifold hypothesis* in machine learning. In what way? Explain the manifold hypothesis, and show mathematically how it relates to the deterministic AE model.

Question 5 (4 points) How many learnable parameters has a 2-layer CNN with bias and max pooling after each convolution (input $\rightarrow$ hidden $\rightarrow$ output), where the input has 100x300 pixels with 20-dimensional features, hidden is composed of 100x300 pixels with 30-dimensional features, and the output has 100x300 pixels and 3-dimensional features? Assume both CNN kernels have size 5x5.

Question 6 (3 points) What is weight sharing? Give one example, and explain how sharing is achieved.

Question 7 (4 points) Consider the translate-and-rotate operator $A_{\theta,v}\, f(x) = f(R_\theta(x+v))$, where $R_\theta = \begin{pmatrix} \sin\theta & \cos\theta \\ -\cos\theta & \sin\theta \end{pmatrix}$. Prove explicitly that it is a linear operator. How would you phrase formally the statement "an operation T is equivariant to translate-and-rotate operations"?

Question 8 (8 points) How would you train an AE so that the latent codes produced by the encoder are as sparse as possible? (i) Give explicit mathematical expressions; (ii) explain your reasoning (iii) discuss about the guarantee (or lack thereof) that at test time you get sparse codes, and (iv) explain why having sparse latent codes might be useful.