Extra task VIR

Elizaveta Isianova

8. 11. 2021

1 Task 1: Understanding of the Activation Functions

Figure 1: Samples from the dataset

Considering having a dataset of pictures of random animals, the task is to classify all animals on a every picture. If you were to design an architecture for a fully connected network, what activation function would you choose and why?
Note that if there are more than one animal of different kinds on a picture, the network has to recognise all of the classes. If there are more than than one animal of the same species, then the network only considers them as one class and does not care about the quantity.

Options:

a. ReLU

b. Sigmoid

c. Softmax

d. Tanh

e. Argmax

Solution:

The correct answer is **Sigmoid.** Neither ReLU, nor Tanh cannot be used for calculating class probability, as the first one returns positive numbers that can be bigger than 1, and the Tanh function returns values in range of (-1; 1). Softmax would work better if there always was only one class on every sample. The same is for Argmax, only it would perform even worse. In this case sigmoid works the best for detecting multiple classes.

2 Task 2: Deriving L2 norm

The next task is to derive the formula of the L2 norm, assuming that the probability of data distribution is given by the Normal distribution: p(y|x∩w) ∼ N(f(x,w),σ^2)

Solution

$$p(y_i|x_i \cap w) = \frac{1}{\sqrt{2\Pi\sigma^2}} * e^{\frac{-(y_i - f(x_i,w))^2}{2\sigma^2}}$$

Consider a loss function in the form of

$$argmin \sum_{i=1}^{n} -log(P(y_i(x_i \cap w))) \Longrightarrow$$

$$\Longrightarrow -log\ p(y_i|x_i \cap w) = -log\frac{1}{\sqrt{2\pi\sigma^2}} * e^{(}\frac{-(y_i - f(x_i,w))}{2\sigma^2}) =$$

$$= -(log\frac{1}{\sqrt{2\pi\sigma^2}} + log\ e^{(}\frac{-(y_i - f(x_i;w))^2}{2\sigma^2})) =$$

$$= -(log\frac{1}{\sqrt{2\pi\sigma^2}} + \frac{(y_i - f(x_i;w))^2}{2\sigma^2}$$

We substitute into the loss function: $\Longrightarrow$

$$\Longrightarrow argmin \sum_{i=1}^{n} -log(\frac{1}{\sqrt{2\pi\sigma^2}}) + \frac{(y_i - f(x_i;w))^2}{2\sigma^2} =$$

$$= argmin \sum_{i=1}^{n} (y_i - f(x_i;w)^2)$$