DeepBayes Summer School Application, 2018

Practical Assignment

Soboleva Natalia, n.sob55@gmail.com

1 Autoencoder architecture

Autoencoder is defined as two parametric differentiable functions, namely the Encoder (E) and the Decoder(D). Parameters of these two functions are adjusted by minimization of the following loss function:

$$L(X, \phi, \theta) = \frac{1}{N} \sum_{i}^{N} \|x_i - D_\phi(E_\theta(x_i))\|_2^2 + \lambda(\|\phi\|_1 + \|\theta\|_1) \to \min_{\phi, \theta}$$

$$X = \{x_i\}_{i=1}^N,\ E_\theta : \mathbb{R}^d \to \mathbb{R}^h,\ D_\phi : \mathbb{R}^h \to \mathbb{R}^d,\ h \ll d,\ \lambda \in \mathbb{R}$$

where the left part is L2 reconstruction loss, the right part is L1 regularizer, and λ is a scalar regularizer weight.

In other word, the main purpose of the following AE is two reduce dimension on the initial data and restore it back with minimal damage as our loss function pay attention on the difference between initial and restored data. This type of AE is usually used in dimension reduction tasks or as noise reducing tools (see).

In current assignment we have a restriction on the number of layers with trainable parameters – six. As our decoder and encoder should consist of the equal number of layers, that means, that we will build AE with 1, 2 or 3 layers in E and D, not including the Dropout layer.

The optimal structure was build according to the describer further experiments.

2 Experiments

2.1 Number of layers and Dropout coefficient

As it was mentioned before, we have only tree options on the number of layers – 1, 2, 3. And as we know, large number of trainable parameters is not too good for the networks, the experiment include adding dropout at the each evaluation. Initially dropout coefficient 0.3 was chosen.

Figure **??** shows Loss/Epoch dependencies on train data and figure **??** on test.

In conclusion, the best result was shown by AE with two encoder layers (with Dropout layer) + two decoder layers – four in total.

2.2 Activation function

In this case on the current AE structure ReLU and tanh were tested as activation functions (see figure **??**).

2.3 Regularization coefficient