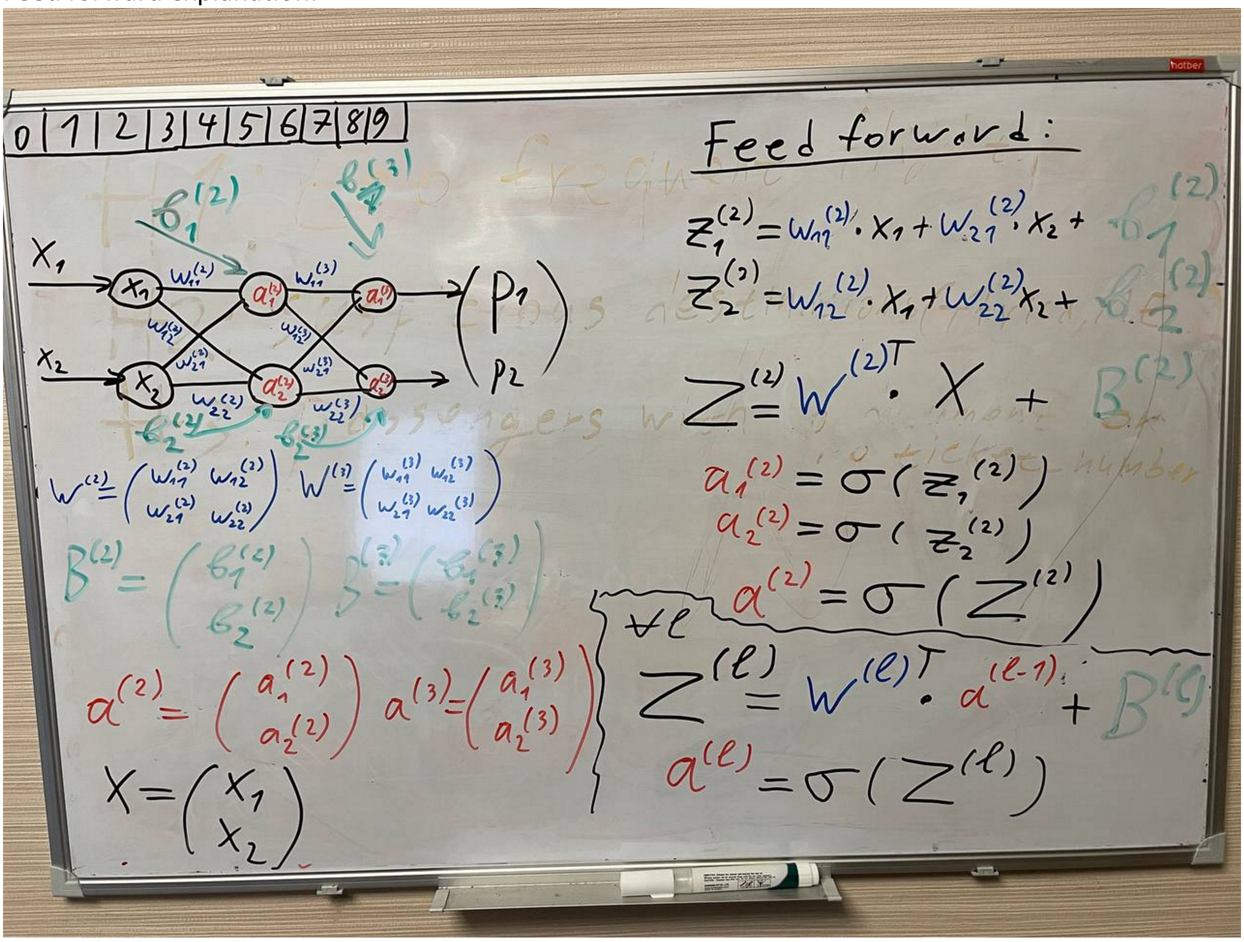
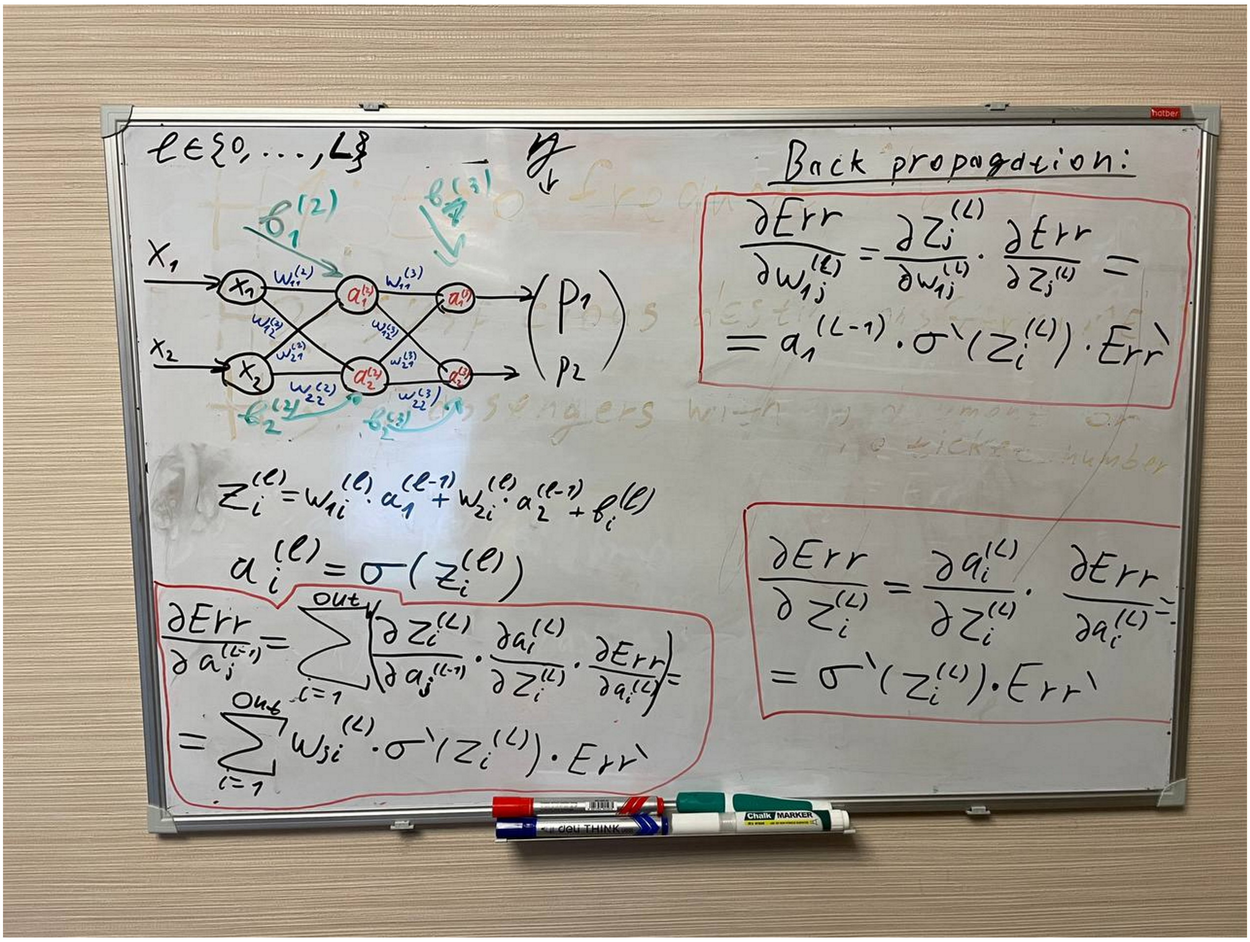
Notes on Neural Networks

19 декабря 2024 г. 10:44

1. Overfitting is usually caused by too high weights. Cause we pay too much attention to some not important inputs
2. How can we deal with that?
3. First approach L1/L2 regularization or Lasso/Ridge. We are penalting the model for having too large weights by adding sum of squares of weights. And multiplying it by alpha to adjust the amount of penalty.
4. Overfitting - validation loss starts rising while training loss still drops. So another startegy is to just early stop. When validation loss is at the minimum
5. Data augmentation - just adjusting training set, so it become more complicated. Especially good for images
6. Drop out - on each iteration random subset of neurons is used. Each neuron is given a hyperparameter p, which stands for probability of neuron being active. So if it is inactive, simply its output is set to zero.
7. Prediction is value of neuron
8. Why not just logistic regression? Because neural networks can extract hidden features, non-linear relationships(when classes can't be separated by a linear function) and be better at handling multi-dimensional data
9. Feed forward explanation:
10. 

11. Backpropagation explanation:

12.



$\ell \in \{0, \dots, L\}$
Back propagation:
$\frac{\partial Err}{\partial w_{1j}^{(\ell)}} = \frac{\partial z_j^{(L)}}{\partial w_{1j}^{(L)}} \cdot \frac{\partial Err}{\partial z_j^{(L)}} =$
$= a_1^{(L-1)} \cdot \sigma'(z_i^{(L)}) \cdot Err'$
$z_i^{(\ell)} = w_{1i}^{(\ell)} \cdot a_1^{(\ell-1)} + w_{2i}^{(\ell)} \cdot a_2^{(\ell-1)} + b_i^{(\ell)}$
$a_i^{(\ell)} = \sigma(z_i^{(\ell)})$
$\frac{\partial Err}{\partial z_i^{(L)}} = \frac{\partial a_i^{(L)}}{\partial z_i^{(L)}} \cdot \frac{\partial Err}{\partial a_i^{(L)}} =$
$= \sigma'(z_i^{(L)}) \cdot Err'$
$\frac{\partial Err}{\partial a_j^{(L-1)}} = \sum_{i=1}^{out} \frac{\partial z_i^{(L)}}{\partial a_j^{(L-1)}} \cdot \frac{\partial a_i^{(L)}}{\partial z_i^{(L)}} \cdot \frac{\partial Err}{\partial a_i^{(L)}} =$
$= \sum_{i=1}^{out} w_{ji}^{(L)} \cdot \sigma'(z_i^{(L)}) \cdot Err'$