Theoretical Part of the Final Exam

Option 1

1) Which of the following is not a regularization tool? (2 points)
    a. Dropout
    b. Shell algorithm
    c. L2
    d. L1
    e. Euclidian Distance
2) What parameter is present in affine layer? (2 points)
    a. Scaling coefficient
    b. Shifting coefficient
    c. Mean variance
    d. Bias
    e. Standard deviation
3) SVM loss is often also called: (2 points)
    a. Marginal loss
    b. Incremental loss
    c. Hinge loss
    d. Decremental loss
    e. Softmax loss
4) The loss optimization algorithm is also called: (2 points)
    a. Forward pass
    b. Normalization
    c. Batch normalization
    d. Layer normalization
    e. Gradient descent
5) The primary purpose of activation functions is to: (2 points)
    a. Make forward pass faster
    b. Make backward pass faster
    c. Generate nonlinearity between the layers of deep neural network
    d. Make deep neural network more flexible
6) What is L2 regularization? Explain how it works on an example. (5 points)
7) Give two reasons to make use of batch normalization with full explanation (5 points)
8) A convolutional neural network has 4 consecutive 3x3 convolutional layers with stride 1 and no pooling. How large is the support of (the set of image pixels which activate) a neuron in the 4th non-image layer of this network? (5 points)

Option 2

1) The raise of convolutional neural networks as we know them today started with: (2 points)
    a. VGG Net
    b. GoogLeNet
    c. ResNet
    d. AlexNet
2) Parameters of neural network that gets tailored by the engineer of the neural network rather than neural network itself are called: (2 points)
    a. Learning parameters
    b. Constant parameters
    c. Hyperparameters
    d. Static parameters
    e. Dynamic parameters
3) Momentum algorithm is the modification of: (2 points)
    a. Forward pass
    b. Gradient descent
    c. Scoring function
    d. Graph optimization
4) The loss determination algorithm is also called: (2 points)
    a. Forward pass
    b. Backward pass
    c. Gradient descent
    d. Scoring
5) Ensemble of models is: (2 points)
    a. Optimization method
    b. A set of neural networks with different architectures performing the same task
    c. A specific neural network type
    d. No such term exists
6) What is L1 regularization? Explain how it works on an example. (5 points)
7) Give two reasons to make use Nesterov gradient descent method with full explanation (5 points)
8) A convolutional neural network has 4 consecutive 3x3 convolutional layers with stride 1 and no pooling. How large is the support of (the set of image pixels which activate) a neuron in the 4th non-image layer of this network? (5 points)