CS461 – RECITATION 13

MACHINE LEARNING PRINCIPLES

Daize Dong

2025-12-08

QUIZ 06

A	B	C	D
-	-	-	-

1.1 Which of these is an advantage of deep neural networks as they are trained today?

A Strong guarantees about convergence to a global minima of the loss function.

B Robust generalization to unseen data.

C The universal approximation theorem means that nonlinear activation functions will be equivalent.

D None of the above are advantages of DNNs.

QUIZ 06

A	B	C	D
-	-	-	-

1.2 Which of these is not a regularization technique used in DNNs?

A Residual layers.

B A large learning rate.

C Randomly zeroing the outputs of some convolutions during training. dropout

D All of the above are regularization techniques.

A large learning rate helps model to jump out of local minimum so it can be treated as a regularizer.

QUIZ 06

A	B	C	D
-	-	-	-

1.3 Which of these helps make it possible to effectively learn with very deep (100s of layers) neural networks?

A Residual layers. ResNet

B A small learning rate.

C Randomly zeroing the outputs of some convolutions during training.

D None of the above allow training of very deep neural networks.

QUIZ 06

A	B	C	D
-	-	-	-

1.4 What can be used as a solution to the vanishing gradient problem?

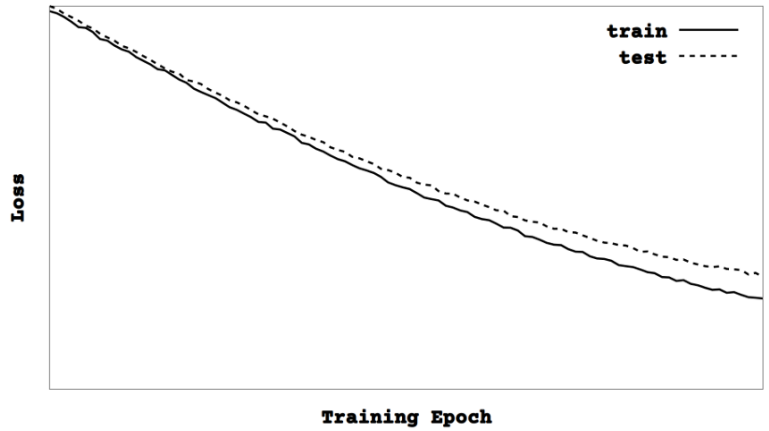
A Train a smaller version of the network, then insert additional intermediate layers after the small network converges.

B Use batch normalization.

C Limit the depth of a network, using wider layers instead of going deeper.

D Any of the above could avoid vanishing gradients.

Batch Norm: keeps activations in a stable range so that they are not too small.



QUIZ 06

A	B	C	D
-	-	-	-

2. Which of the following statements seem to be true about the DNN being trained given the loss curve in this graph?

A There is some mismatch between the training data and the test data.

B There is a mismatch between the training and testing data, but is probably unbiased noise.

C The training data must be easier than the testing data.

D Even if the training loss reaches zero, the testing loss will not converge to zero.

QUIZ 06

A	B	C	D
–	–	–	–

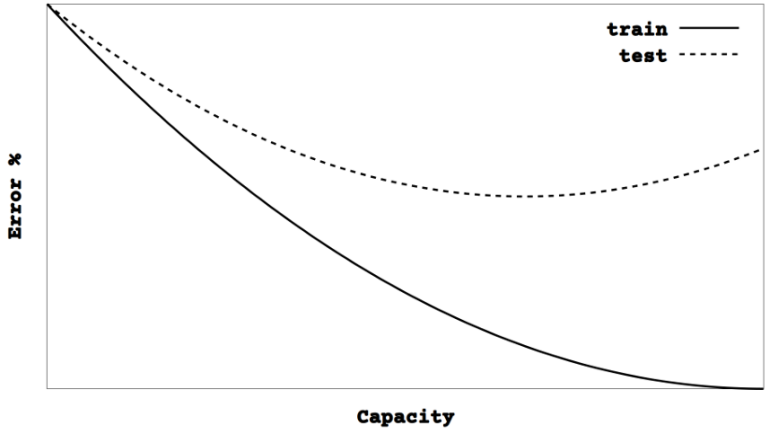
3. You add data augmentation to the training set in an attempt to improve results. The previous loss curves are present in gray, and the new training loss curve has higher loss while the testing loss curve has lower loss. Which of the following correctly describe what could be happening and how you should interpret the results?

A The mismatch between the training data and the test data is greater, and this augmentation is bad.

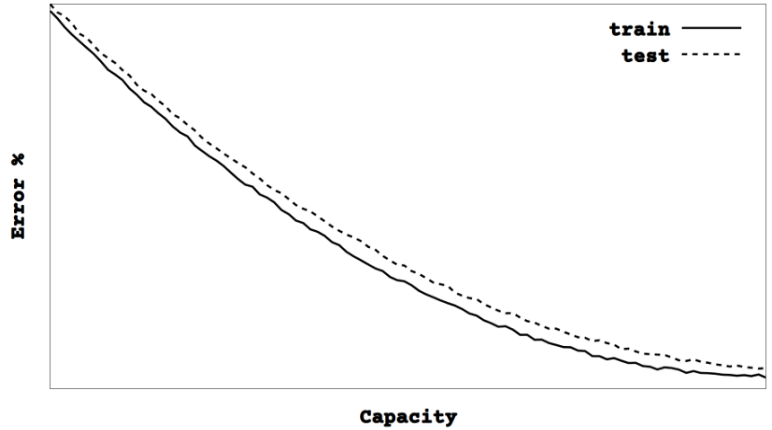
B The testing loss is lower, so this augmentation is good.

C The augmentation has made the training set more difficult to learn.

D This augmentation is acting as a regularizer by getting the DNN to generalize better on unseen data.



QUIZ 06



4. You begin training a *linear* neural network. You increase its capacity by adding more fully connected units in the hidden layers, and observe the curve on the left. The next day, you decide to switch to a *convolutional neural network* that ends with a small linear network that serves as a classifier. You start small, but add capacity by increasing the number of kernels and feature maps and observe the figure on the right. What can you say about these two networks and your dataset?

A The fully linear network is more prone to overfitting than the convolutional network.

B There are examples in the test set that do not appear in the training set.

C Convolutional networks are more powerful than linear networks, and will get lower error rates regardless of the differences in training and test sets.

D The convolutional results look better at first glance, but since the error rate on the training set has gone up they are actually a bad choice for this dataset.

QUIZ 06

5. What is the receptive field of two stacked 3x3 convolutions?

5x5

6. A 5x5 convolution is run over a 32x32 pixel image. There is no padding and the stride is 1. What are the dimensions of the output feature map?

28x28