Deep Learning Mentimeter

In DL if you minimize a loss function f(x). What is x?

- The learnable parameters

How can you distinguish a minimum from a maximum or saddle point in 1D?

- By looking at the second derivative

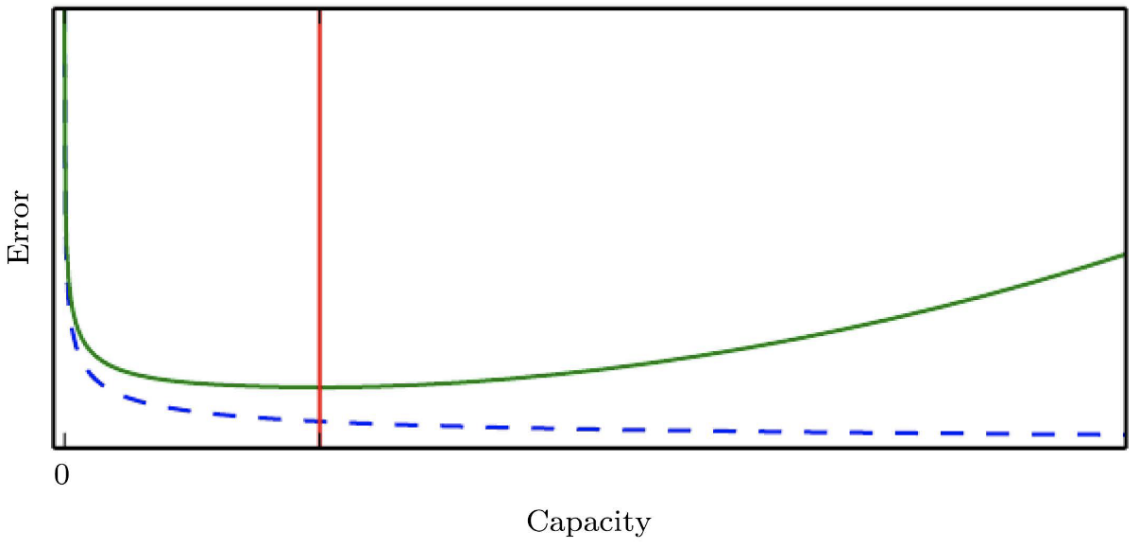
How can you distinguish a minimum from a maximum or saddle point in multidimensional problems?

- Eigenvalues of the Hessian

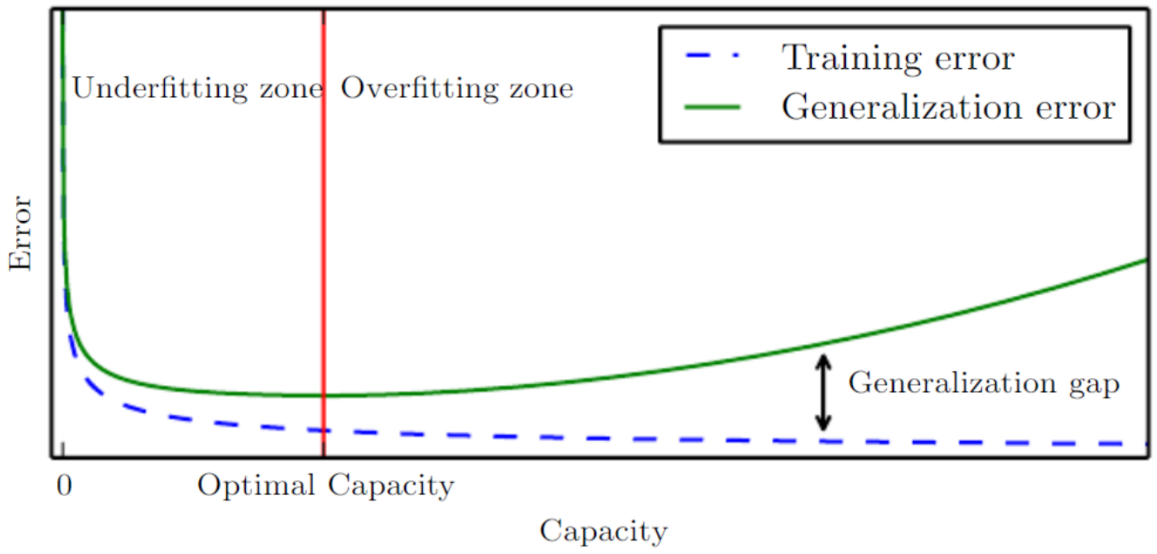
What is part of the i.i.d. assumptions?

- independent, identically distributed

What is the blue curve in the plot?



- The training error



What is the goal of regularization?

- Reduce generalization gap

Which statement about maximum likelihood (ML) estimation is correct?

- ML is equivalent to minimizing the cross-entropy!

Which statement about SGD (stochastic gradient descent) is correct?

- The variance of SGD is larger than the variance of gradient descent

Which statement about manifolds is correct?

- A road is a 1d manifold in our 3d world

What is the following function g(x) in the following hidden layer called?

$$h = g(W^\top x + b)$$

- Activation function

Which activation function do you choose for the hidden layers in a classification neural network?

- ReLu Function

Which output unit do you choose for a regression neural network?

- Linear

Which output unit do you choose for a multi-class classification neural network?

- Softmax ⇒ pmf

What is calculated using backpropagation?

- Gradient

Over what do we iterate in backpropagation?

- All layers of the neural network

In a neural network with 1000 weights, what is the size of the gradient?

- 1000

What is the goal of regularization?

- Reduce test error

In which direction does weight decay (L2) regularize strongly?

- Directions with small curvature

What is the idea of dataset augmentation?

- Increase dataset size to improve generalization

What is the idea of early stopping?

- Stop when the validation loss gets larger

Which statement about ensemble methods is true?

- In ensemble you average the output of multiple models

What do you do in dropout?

- Randomly drop nodes during each iteration of training

What do you need to do when applying dropout after training?

- Rescale the weights

What is a training algorithm that uses the entire training data in each update step called?

- Batch method

How is the mini-batch size in mini-batch stochastic gradient descent selected?

- Pick based on the available GPU memory

Do you expect more local minima or saddle points in a neural network?

- Saddle points

What is the idea of Xavier initialization?

- Initialize so that activations and gradients have unit variance

What is the idea of momentum?

- Smooth noisy gradients

What is the idea of batch normalization?

- Normalize the hidden units

True or false: Convolution is computationally more efficient than matrix multiplication?

- True

True or false: Convolution uses less memory than matrix multiplication?

- True

True or false: Pooling makes a CNN invariant to small rotations of the input?

- False

True or false: Convolution usually needs no non-linear activation function?

- False

True or false: Using correlation instead of convolution in a CNN improves performance?

- False