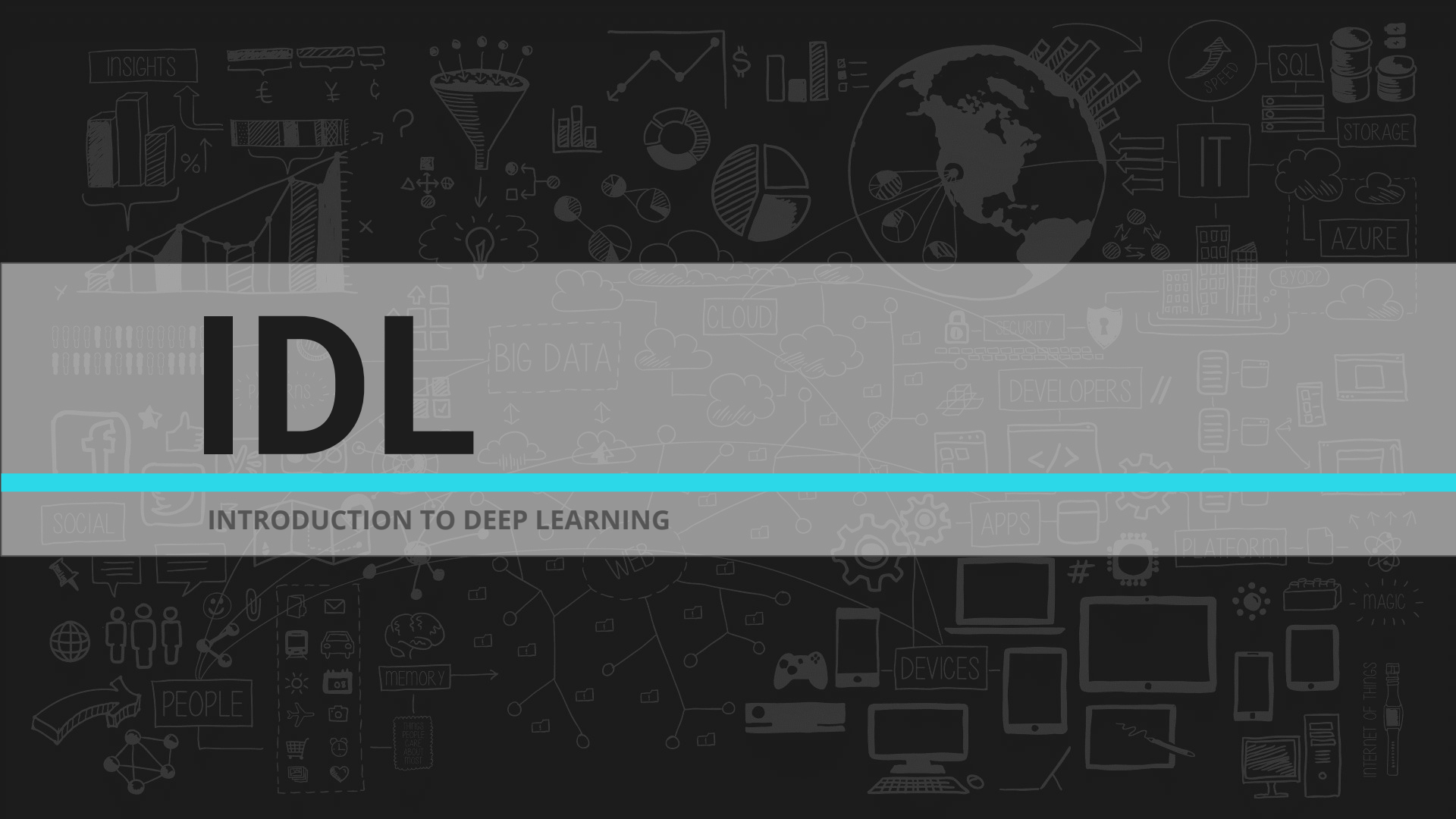
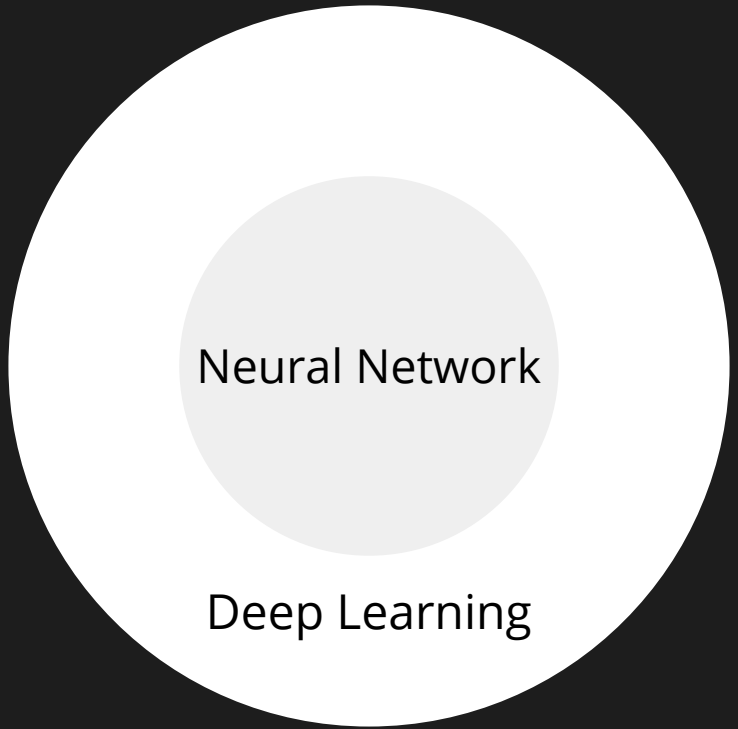


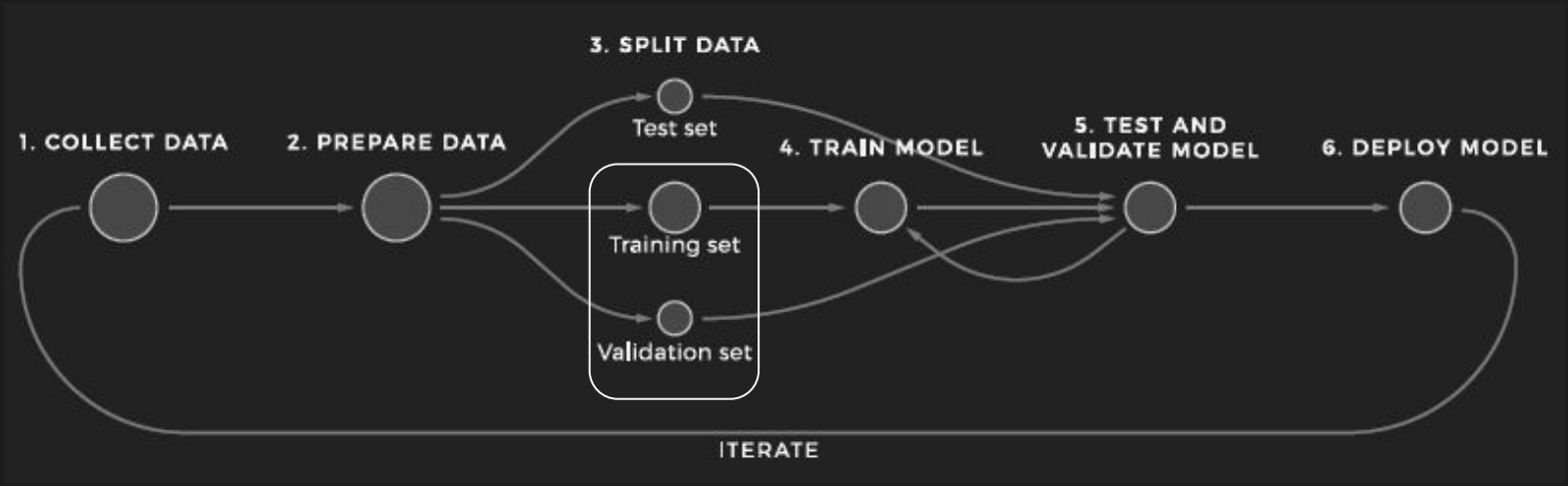
IDL

INTRODUCTION TO DEEP LEARNING

Deep learning



Train-Dev-Test Set



If size of the dataset is 100 to 1000000 ==> 60/20/20
If size of the dataset is 1000000 to INF ==> 98/1/1 or 99.5/0.25/0.25

Bias Variance Tradeoff

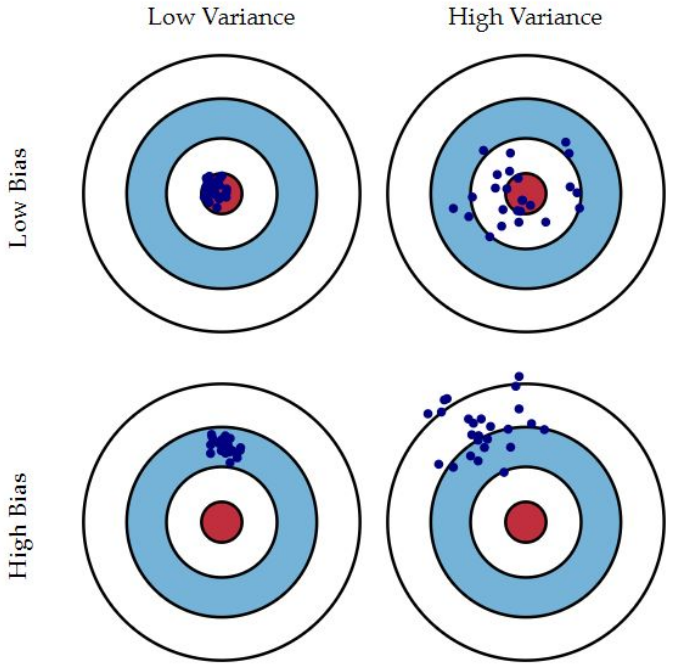
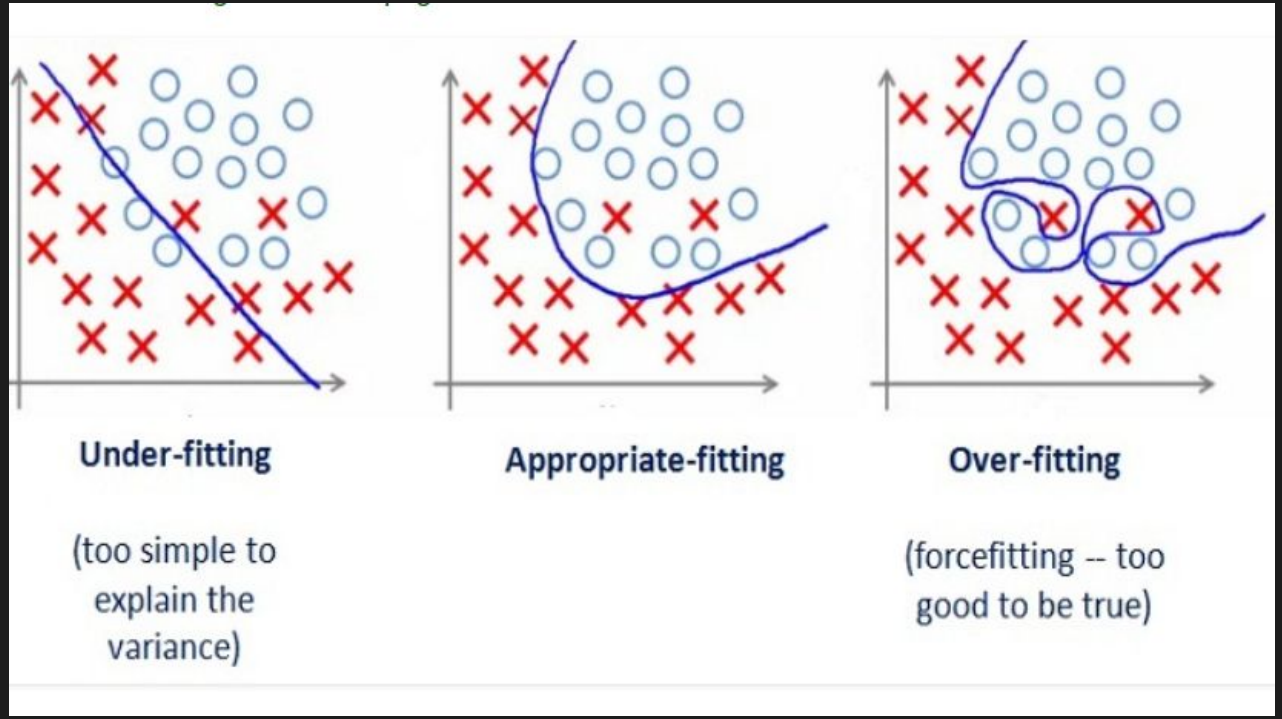
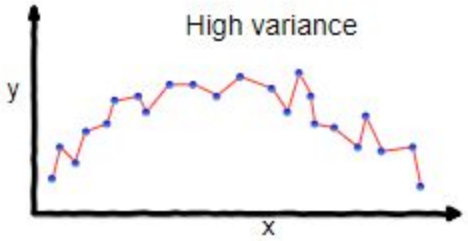


Fig. 1 Graphical illustration of bias and variance.

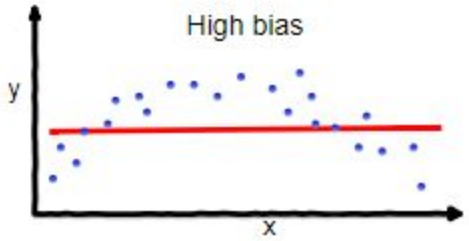
Overfitting and Underfitting



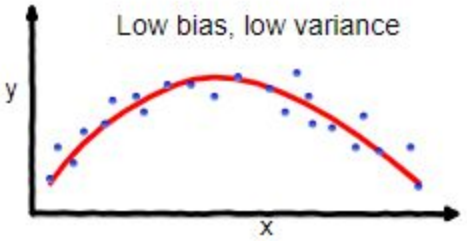
Bias Variance Tradeoff



overfitting



underfitting



Good balance

Overfitting and Underfitting

Train Set :	1%	15%	15%	0.5%
Dev Set :	10%	16%	30%	1.%
	High Variance	High Bias	High Variance & High Bias	Low Variance & Low Bias

Solve High Variance and High Bias

If your algorithm has a high bias:

- Try to make your NN bigger (size of hidden units, number of layers)
- Try a different model that is suitable for your data.
- Try to run it longer.
- Different (advanced) optimization algorithms.

If your algorithm has a high variance:

- More data.
- Try regularization.
- Try a different model that is suitable for your data.

Normalizing inputs

- speed up the training process
- optimization will be faster

Normalization are going on these steps:

- Get the mean of the training set: mean = (1/m) * sum(x(i))
- Subtract the mean from each input: X = X - mean
  - This makes your inputs centered around 0.
- Get the variance of the training set:
  - variance = (1/m) * sum(x(i)^2)
- Normalize the variance. X /= variance

- training, dev, and testing sets (but using mean and variance of the train set)

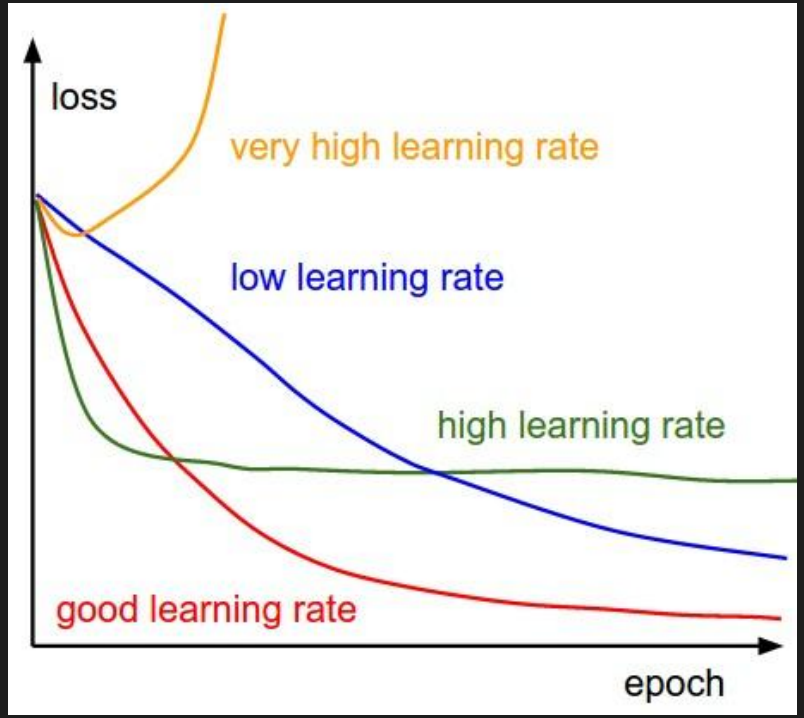
Epochs

an epoch is a single pass through the full training set.

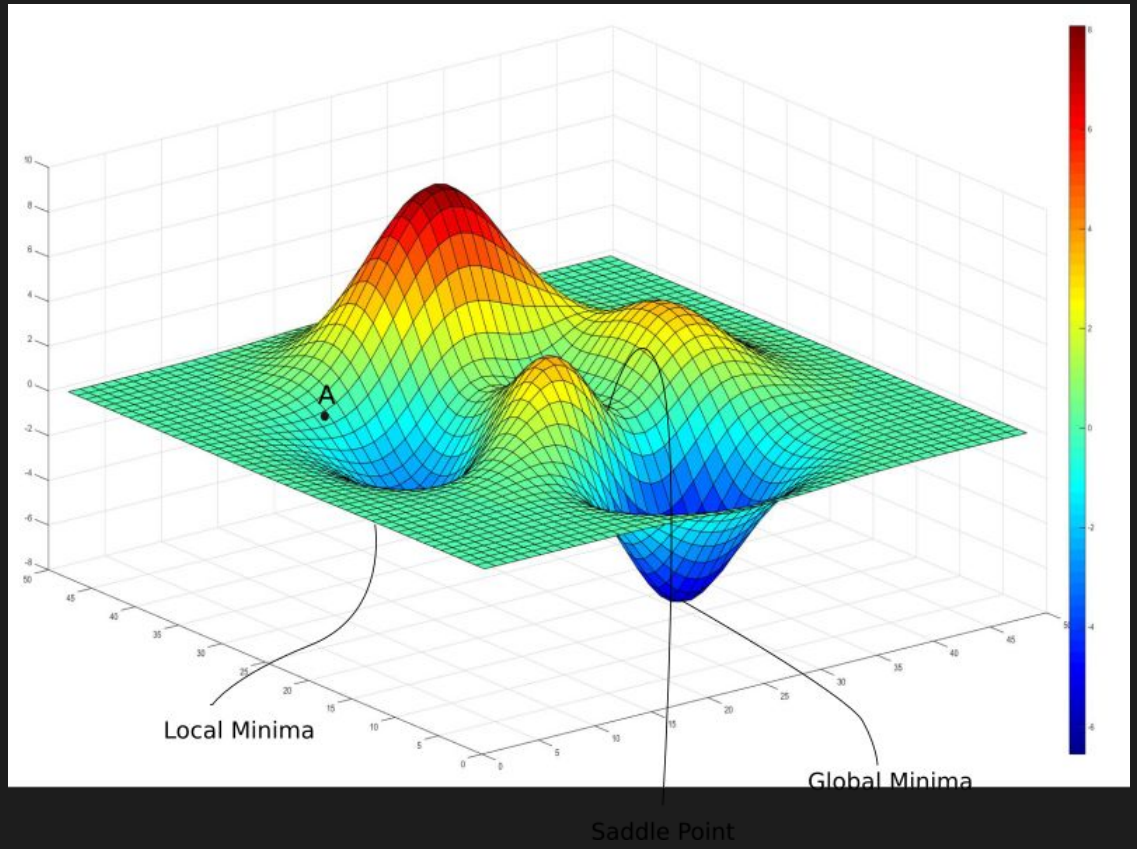
Hyperparameter

Machine algorithms’ settings that must be determined external to the learning algorithm itself

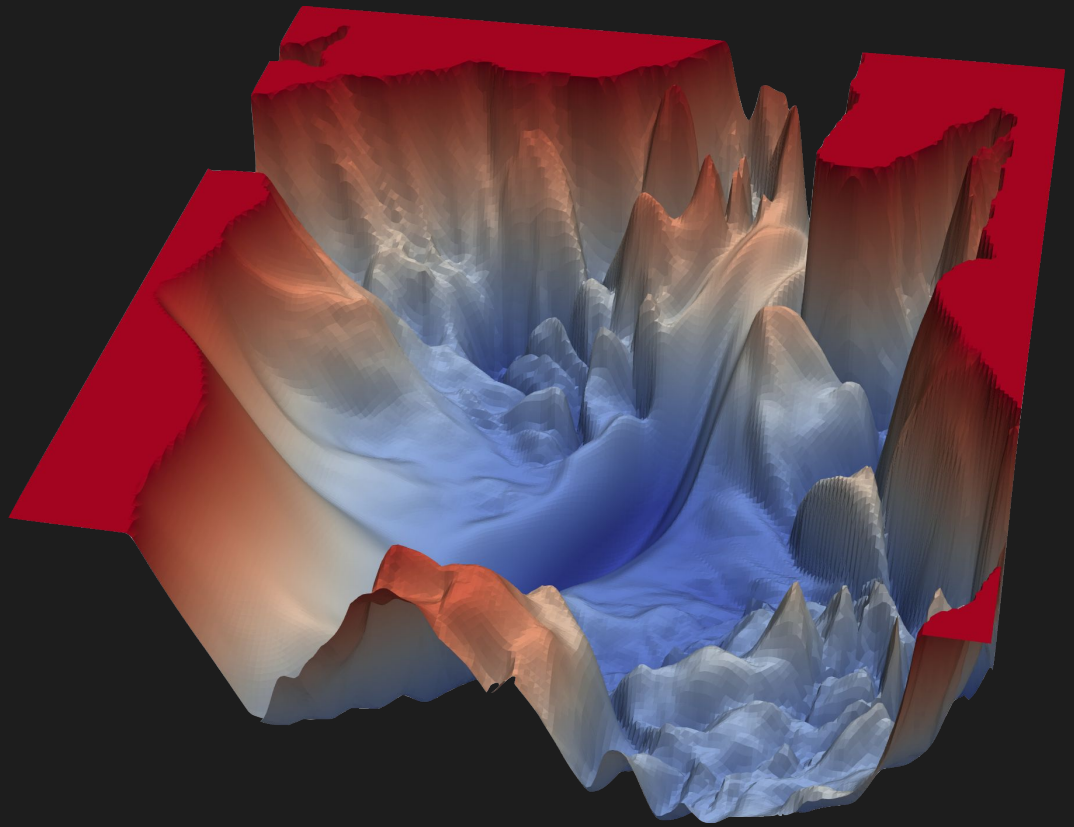
Learning Rate



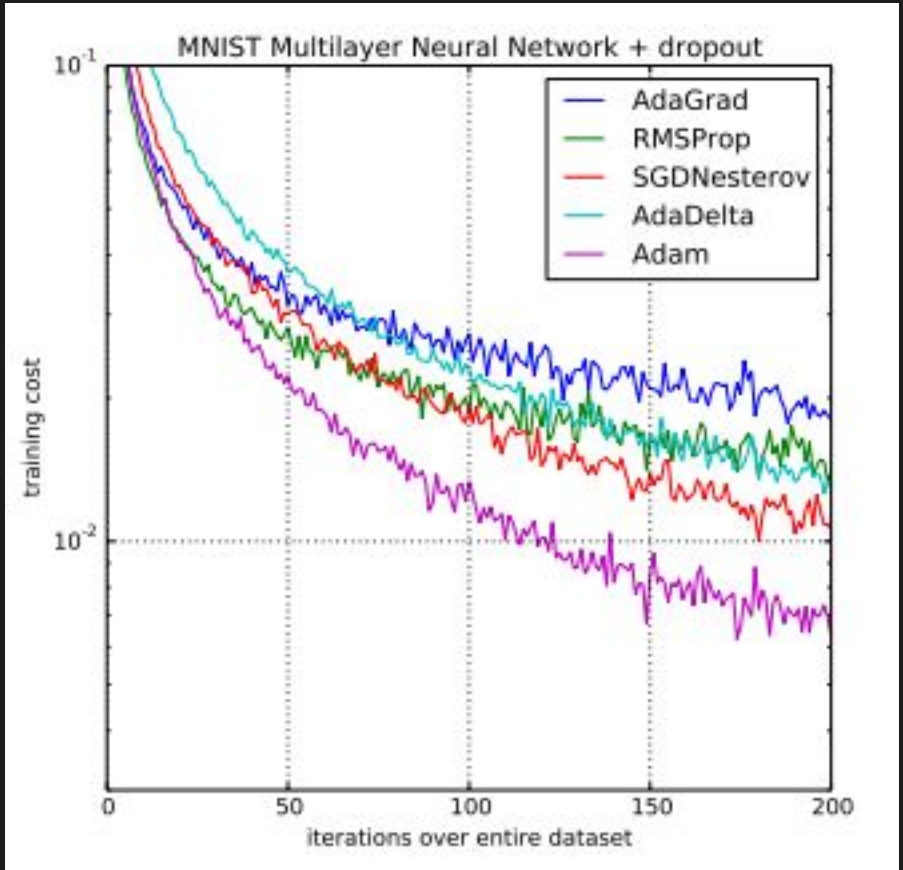
Loss function



Loss function



Optimizer


MNIST Multilayer Neural Network + dropout
AdaGrad
RMSProp
SGDNesterov
AdaDelta
Adam
training cost
10^{-1}
10^{-2}
0
50
100
150
200
iterations over entire dataset

MNIST