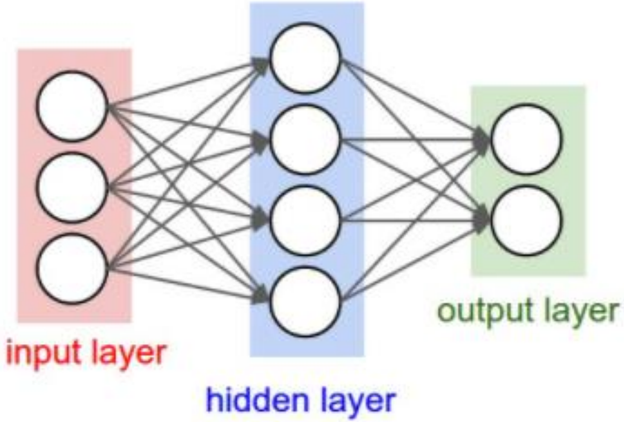
Introduction to Deep Learning (I2DL)

Exercise 5: Neural Networks and CIFAR10 Classification

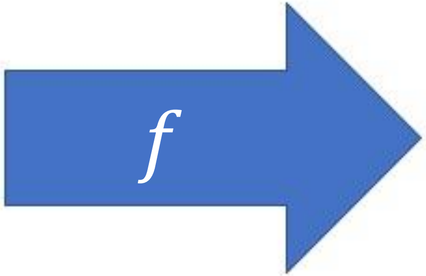
Today's Outline

- Neural Networks
  – Mathematical Motivation
  – Modularization

- Exercise 5
  – Implementation Loop
  – CIFAR10 Classification



Our Goal

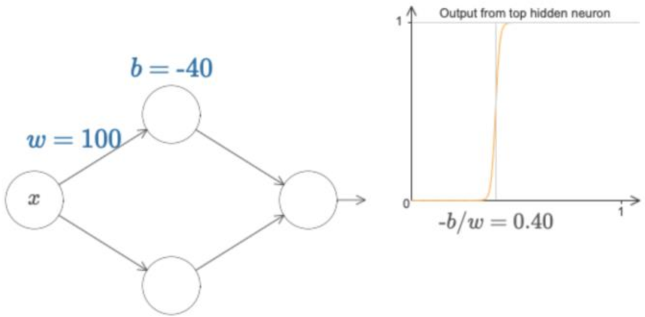


$?

Universal Approximation Theorem

Theorem (1989, colloquial)

For any continuous function f on a compact set K, there exists a one layer neural network, having only a single hidden layer + sigmoid, which uniformly approximates f to within an arbitrary $\varepsilon > 0$ on K.



Universal Approximation Theorem (Optional)

Readable proof:

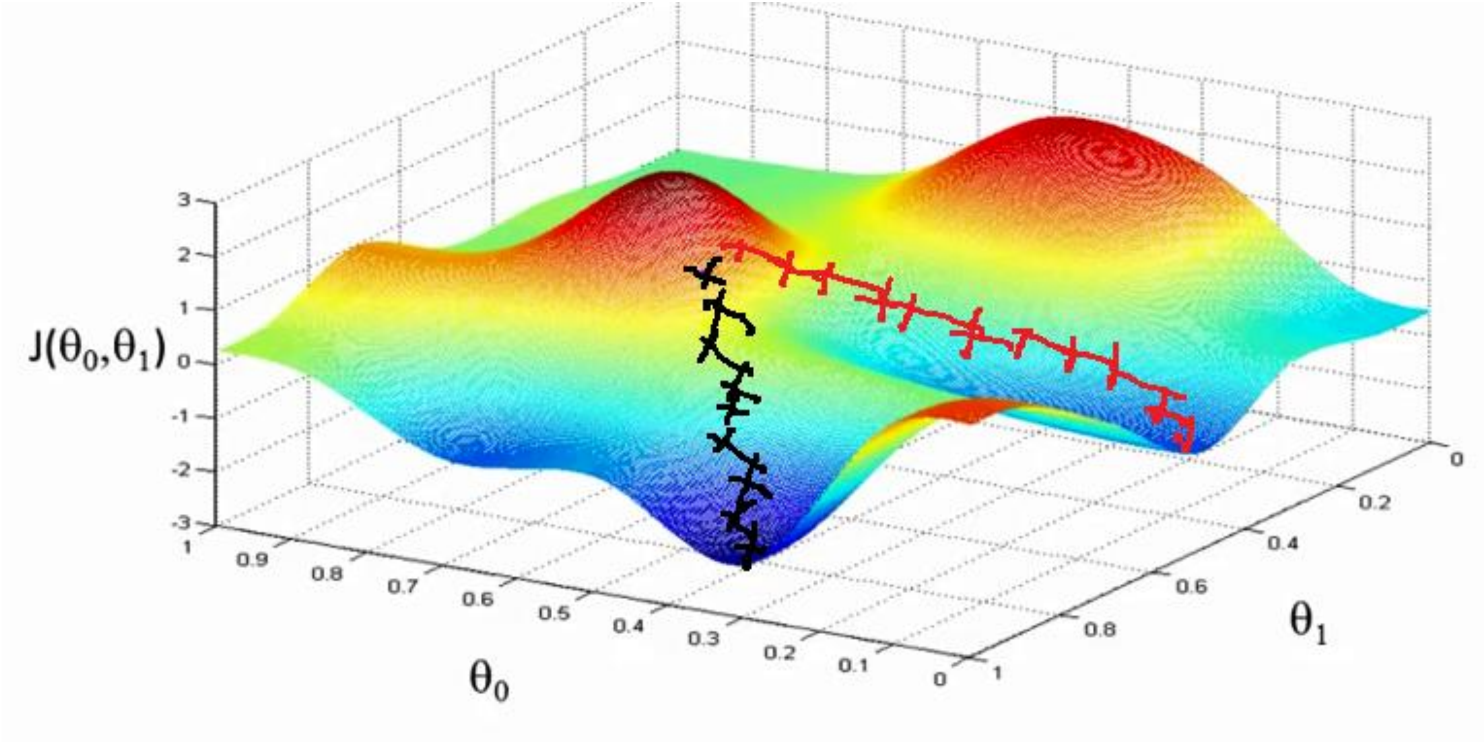
https://mcneela.github.io/machine_learning/2017/03/21/Universal-Approximation-Theorem.html

(Background: Functional Analysis, Math Major 3rd semester)

Visual proof:

http://neuralnetworksanddeeplearning.com/chap4.html

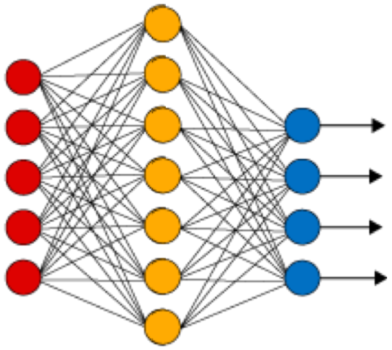
A word of warning…



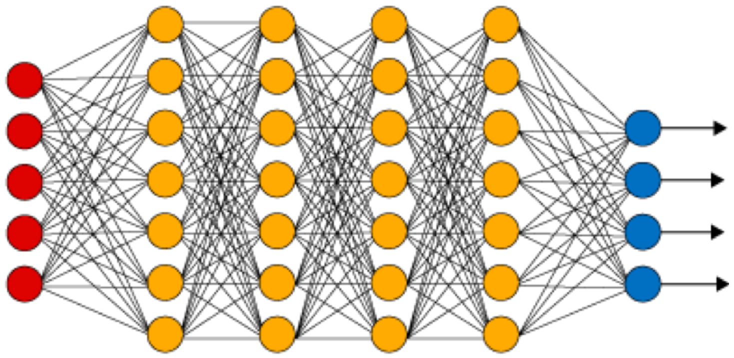
Source: http://blog.datumbox.com/wp-content/uploads/2013/10/gradient-descent.png

Shallow vs. Deep

- Shallow
  (1 hidden layer)



- Deep
  (>1 hidden layer)



Obvious Questions

- Q: Do we even need deep networks?
  A: Yes. Multiple layers allow for more representation
  power given a fixed computational budget
  incomparison to a single layer

- Q: So we just build 100 layer deep networks?
  A: Not trivially ;-)
  Contraints: Memory, vanishing gradients, …

Exercise 4: Simple Classification Net

```python
class Classifier(Network):
    """
    Classifier of the form y = sigmoid(X * W)
    """

    def __init__(self, num_features=2):
        super(Classifier, self).__init__("classifier")

        self.num_features = num_features
        self.W = None

    def initialize_weights(self, weights=None)[illegible]
        """
        Initialize the weight matrix W

        :param weights: optional weights for i[illegible]al[illegible]tion
        """
        if weights is not None:
            assert weights.shape == (self.num_features + 1, 1), \
                "weights for initialization are not in the correct sh[illegible]e
            self.W = weights
        else:
            self.W = 0.001 * np.random.randn(self.num_featu[illegible]1,
```

```python
def forward[illegible]lf,[illegible]
    """
    Performs the forwa[illegible] of the model.

    [illegible]aram [illegible]N x D array of training data. Each row is a D-dimensional point.
    [illegible]turn: [illegible]dicted labels for the data in X, shape N x 1
             1-dimensional array of length N with classification scores.
    """
    assert self.W is not None, "weight matrix W is not initialized"
    # add a column of 1s to the data for the bias term
    batch_size, _ = X.shape
    X = np.concatenate((X, np.ones((batch_size, 1))), axis=1)
    # save the samples for the backward pass
    self.cache = X
    # out[illegible]ble
    y = [illegible]e

    #[illegible]#####################################################
    # [illegible]O:                                                            #
    # I[illegible]ement the forward pass and return the output of the model. Note  #
    # that you need to implement the function self.sigmoid() for that     #
    ########################################################################

    y = X.dot(self.W)
    y = self.sigmoid(y)

    ########################################################################
    #                           END OF YOUR CODE                          #
    ########################################################################
```

Doesn't scale

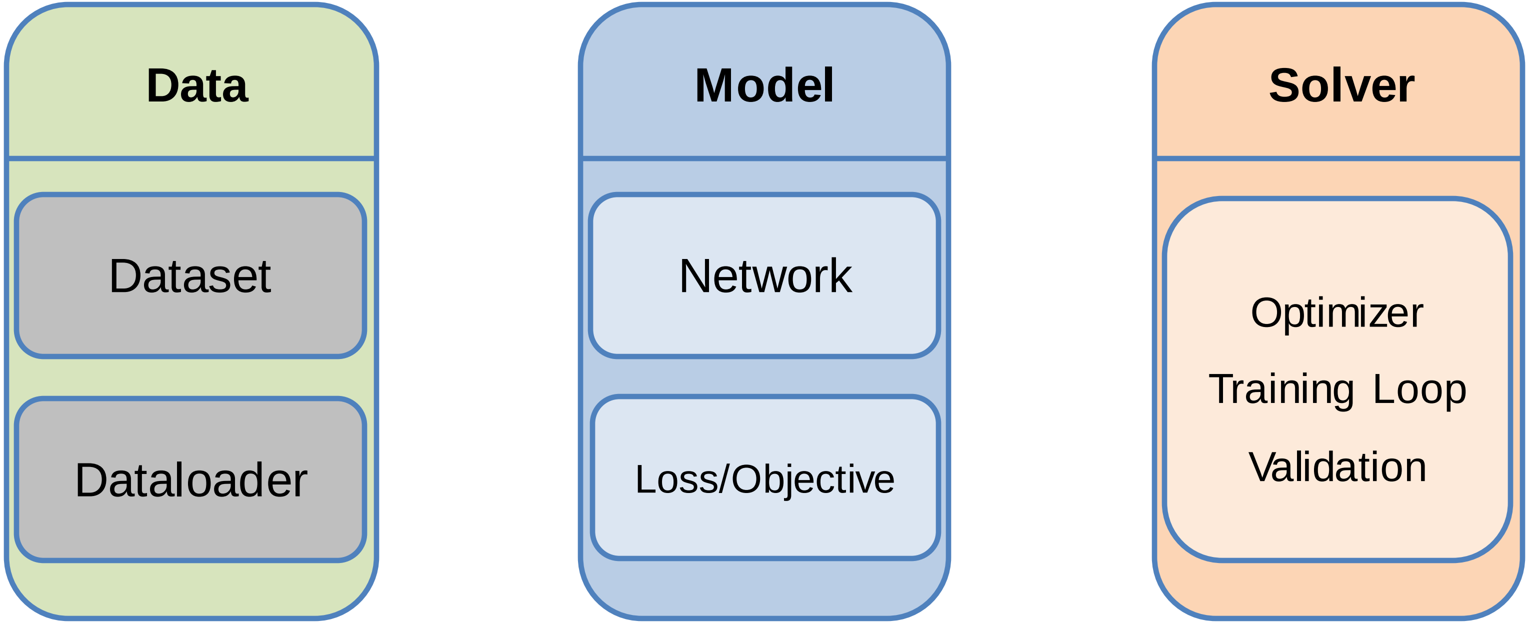
Modularization

Chain Rule:

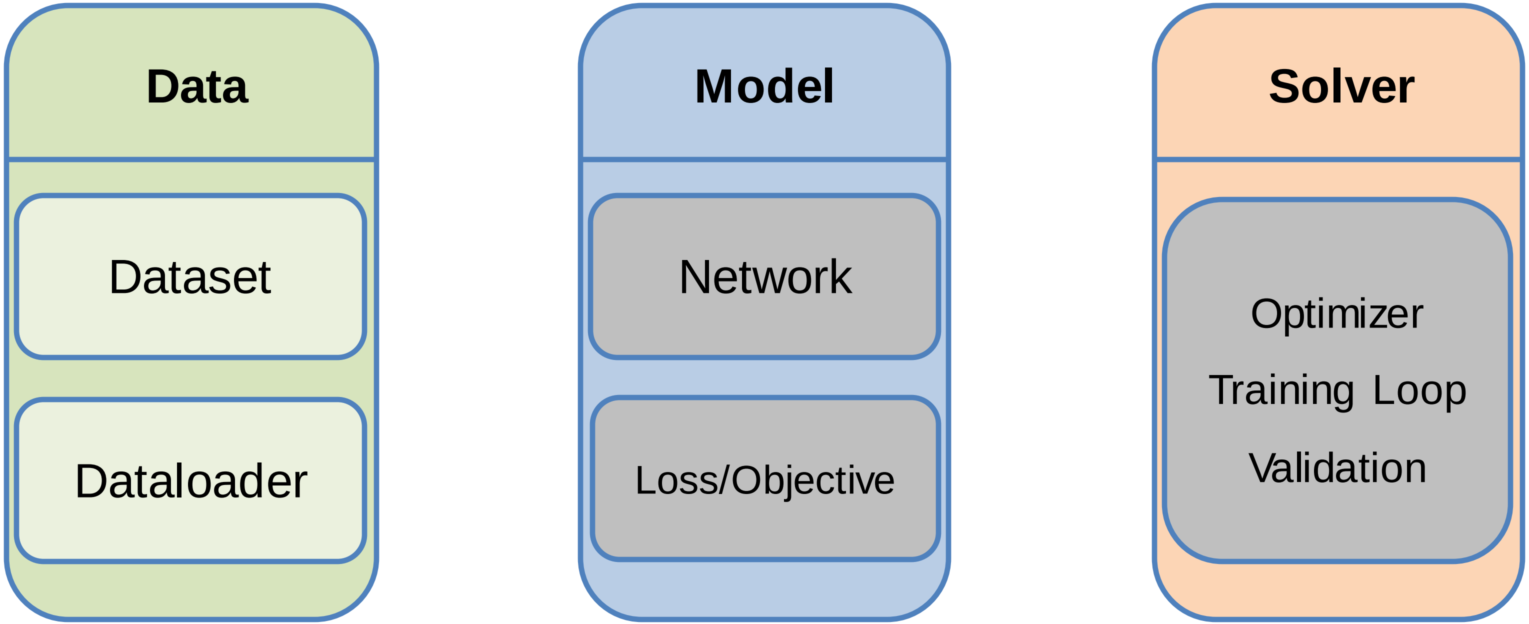
$$\boxed{\frac{\partial f}{\partial y} = \frac{\partial f}{\partial d} \cdot \frac{\partial d}{\partial y}}$$

```python
class Sigmoid:
    def __init__(self):
        pass

    def forward(self, x):
        """
        :param x: Inputs, of any shape

        :return out: Output, of the same shape as x
        :return cache: Cache, for backward computation, of the same shape as x
        """

    def backward(self, dout, cache):
        """
        :return: dx: the gradient w.r.t. input X, of the same shape as X
        """
```

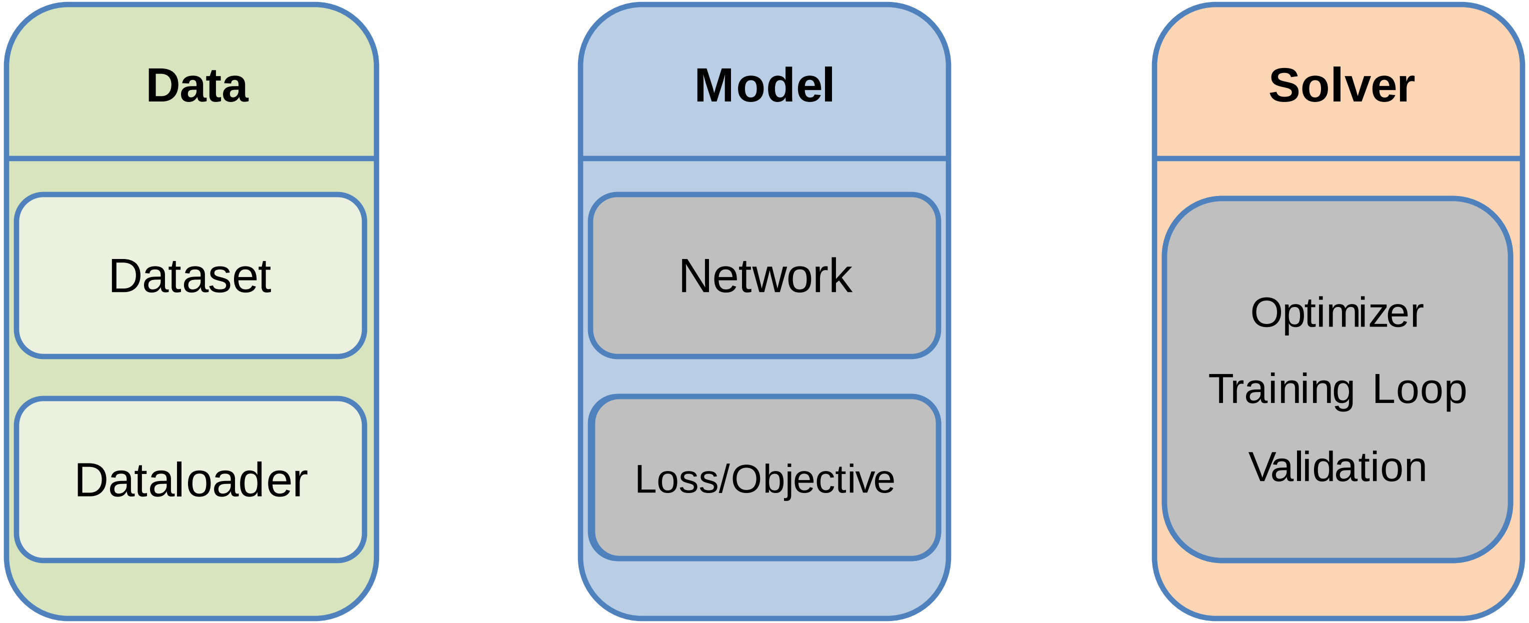
Exercise 3: Dataset



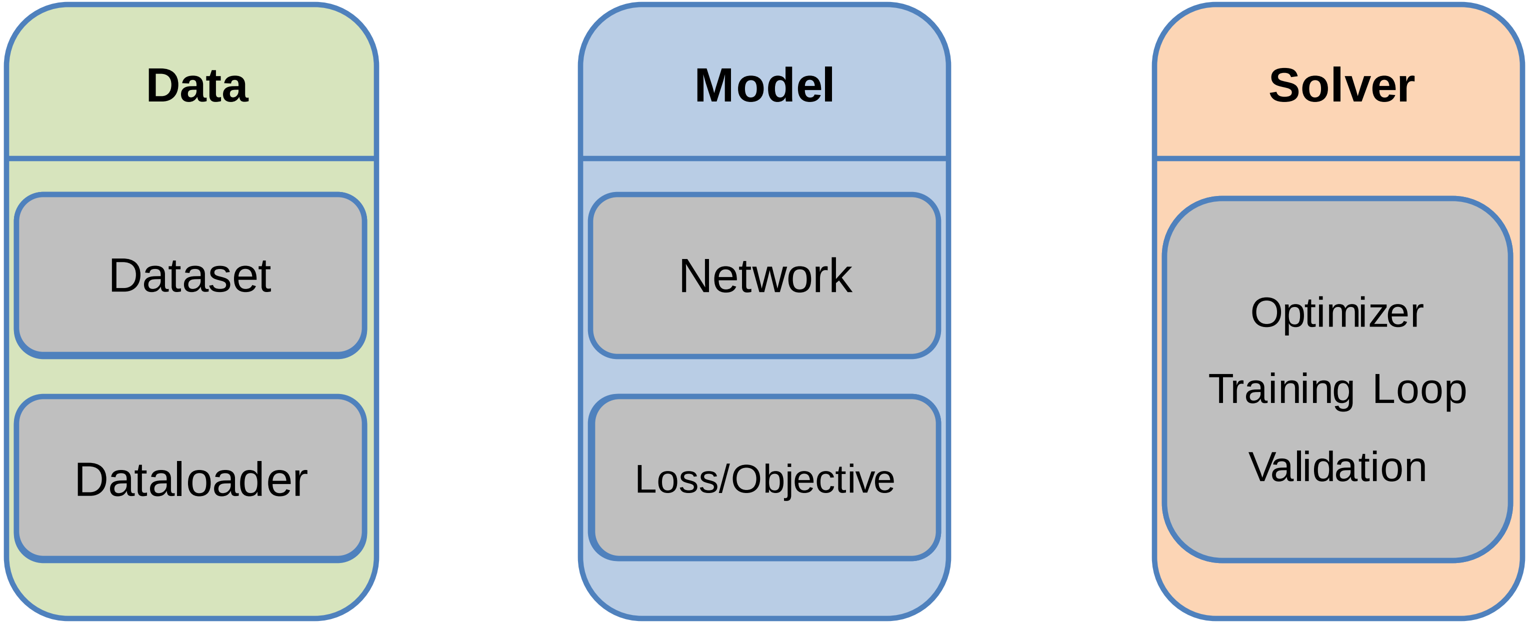
Exercise 4: Binary Classification



Exercise 5: Neural Networks and CIFAR10 Classification



Exercise 6: Neural Networks and Hyperparameter Tuning



Summary

- Monday 22.11: Watch Lecture 6
  - Training NNs

- Wednesday 24.11 15:59: Submit exercise 5

- Thursday 25.11: Tutorial 6
  - Hyper-parameter Tuning

See you next week
☺