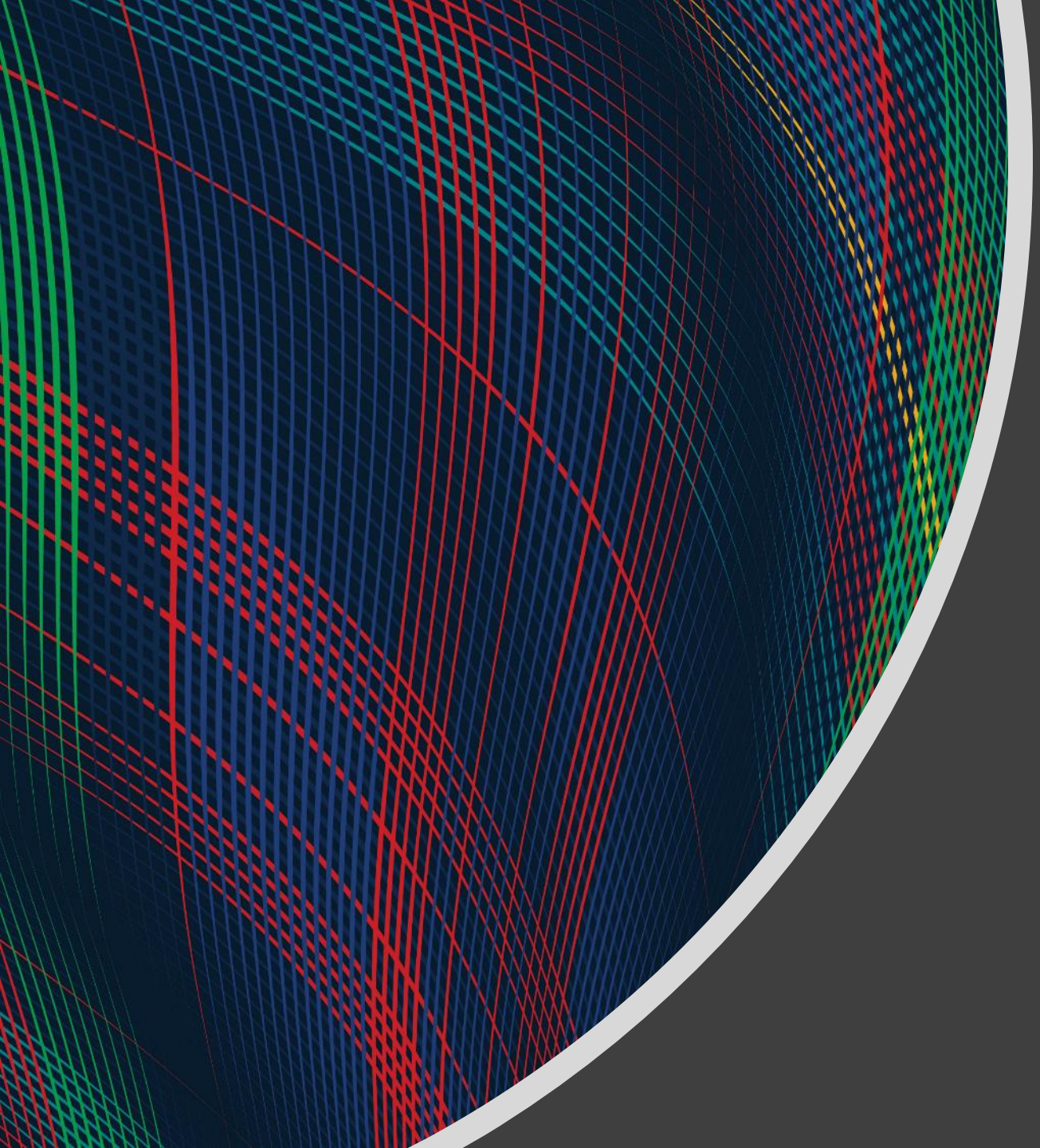


Demystifying AI

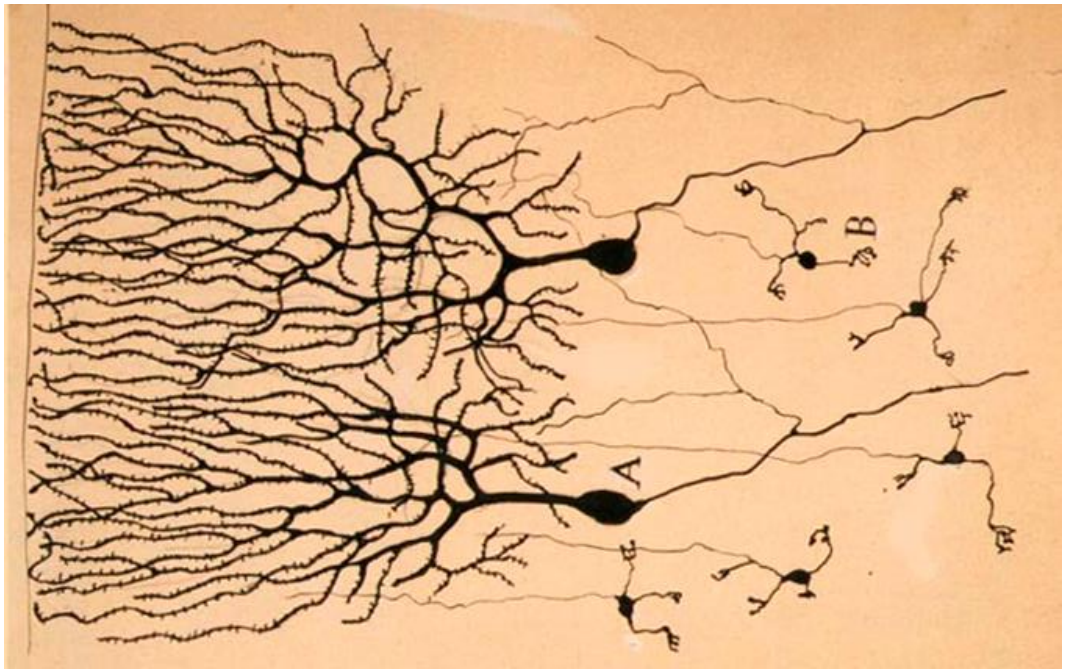
Neural Network Optimization

Instructor: Pat Virtue

Neural Networks

Inspired by actual human brain

Input Signal



Output Signal

DOG

CAT

TREE

CAR

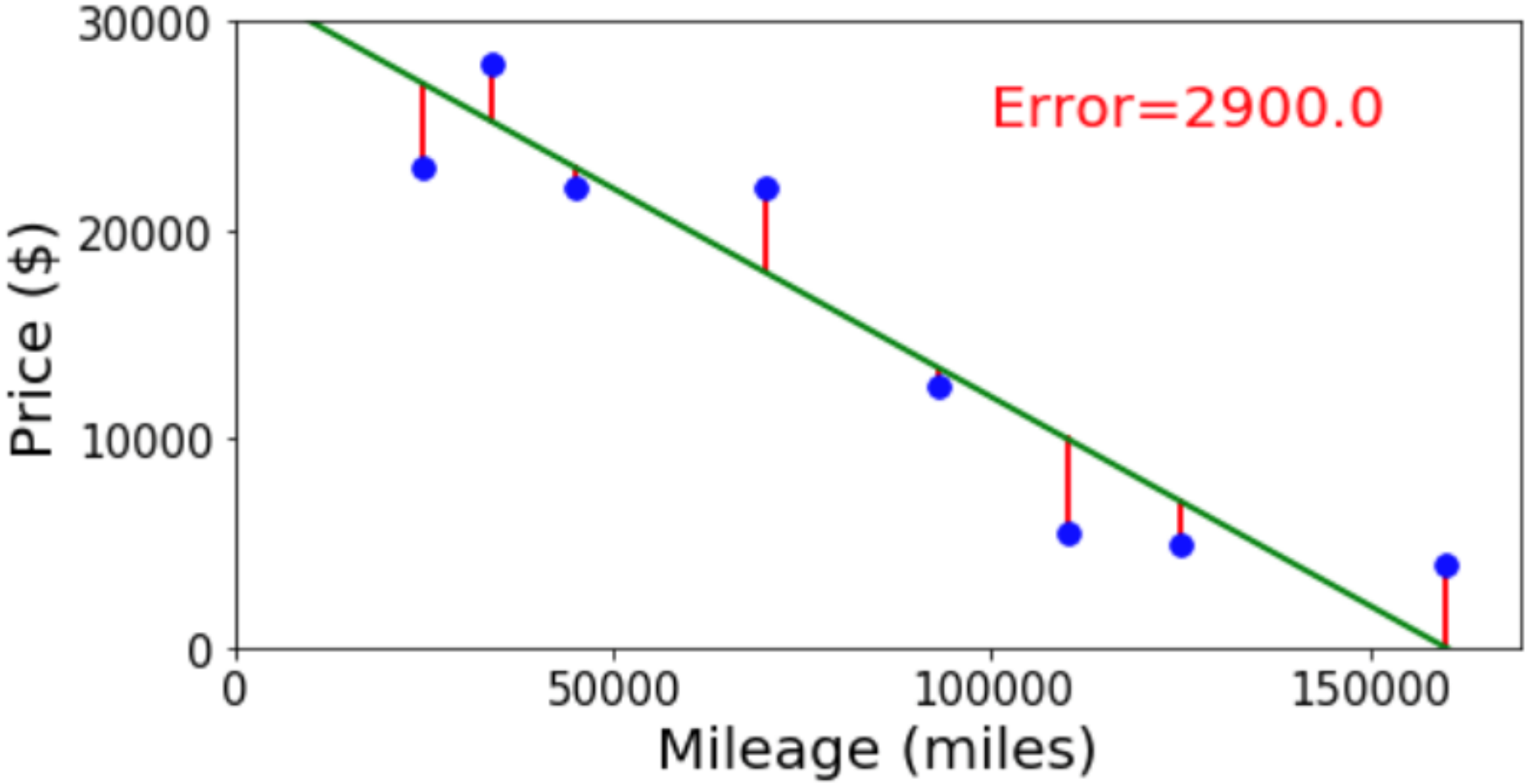
SKY

Image: https://en.wikipedia.org/wiki/Neuron

Neural Networks

Simple single neuron example:

- Selling my car



Neural Networks

Many layers of neurons, millions of parameters

Input Signal

Output Signal

DOG

CAT

TREE

CAR

SKY

Neural Networks

Many layers of neurons, millions of parameters

Input Signal

Output Signal

DOG

CAT

TREE

CAR

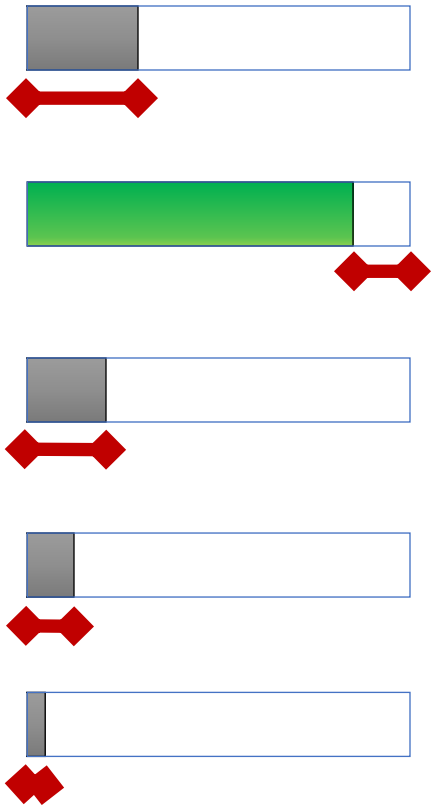
SKY

Neural Networks

Many layers of neurons, millions of parameters

Input Signal

Output Signal



LEFT

RIGHT

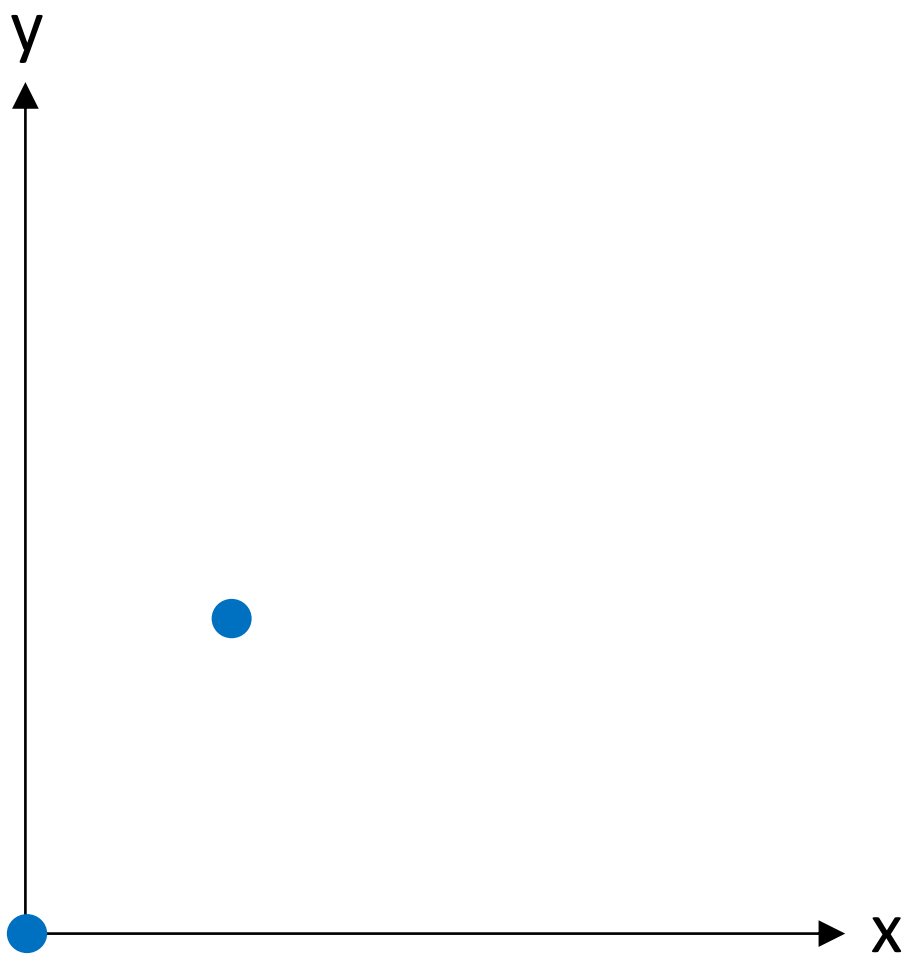
UP

DOWN

BUTTON

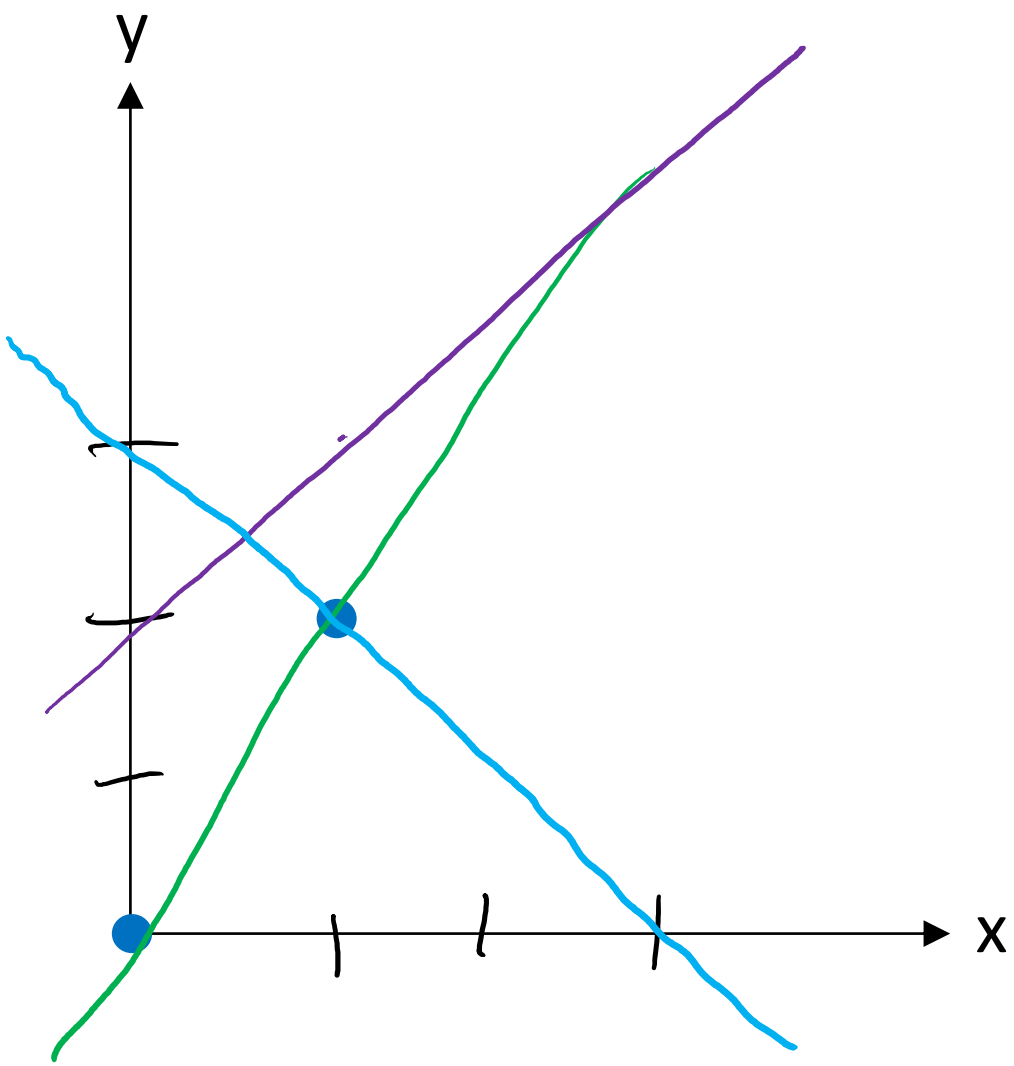
Reminder: Training: Finding the Best Parameters

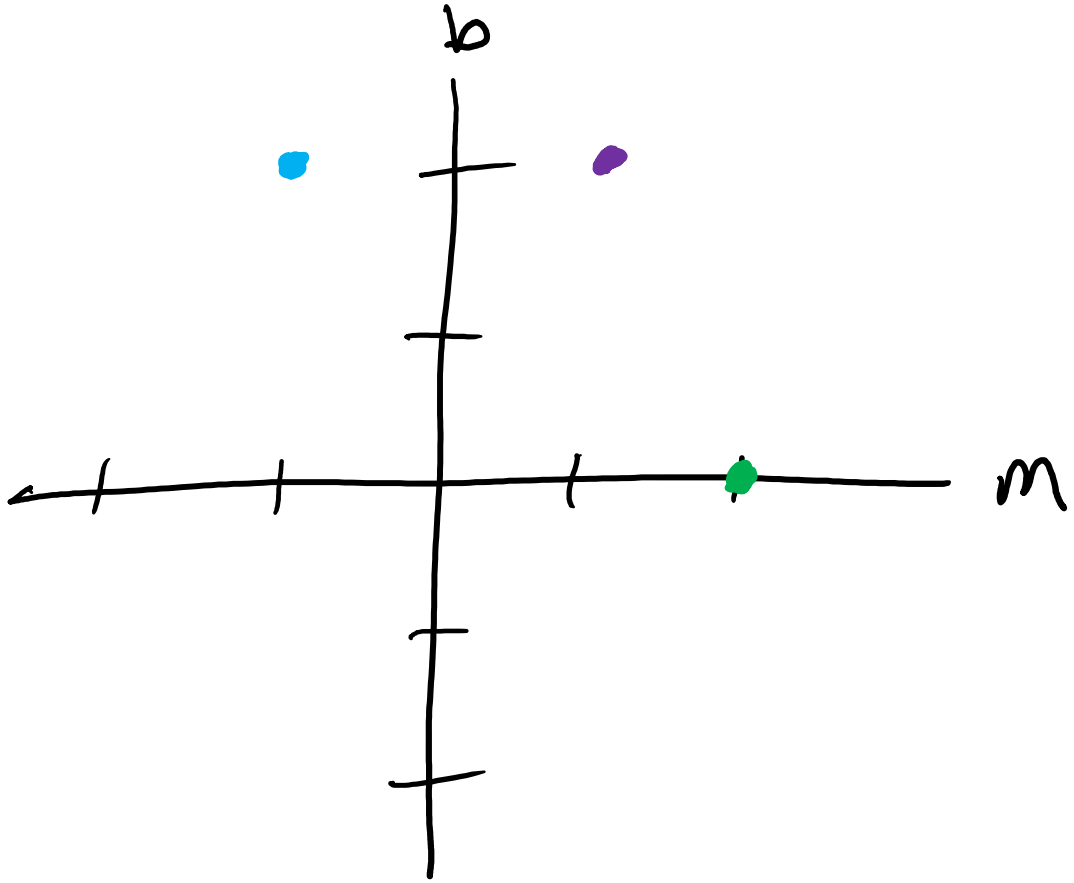
Searching for parameters that minimize mean squared error



Reminder: Training: Finding the Best Parameters

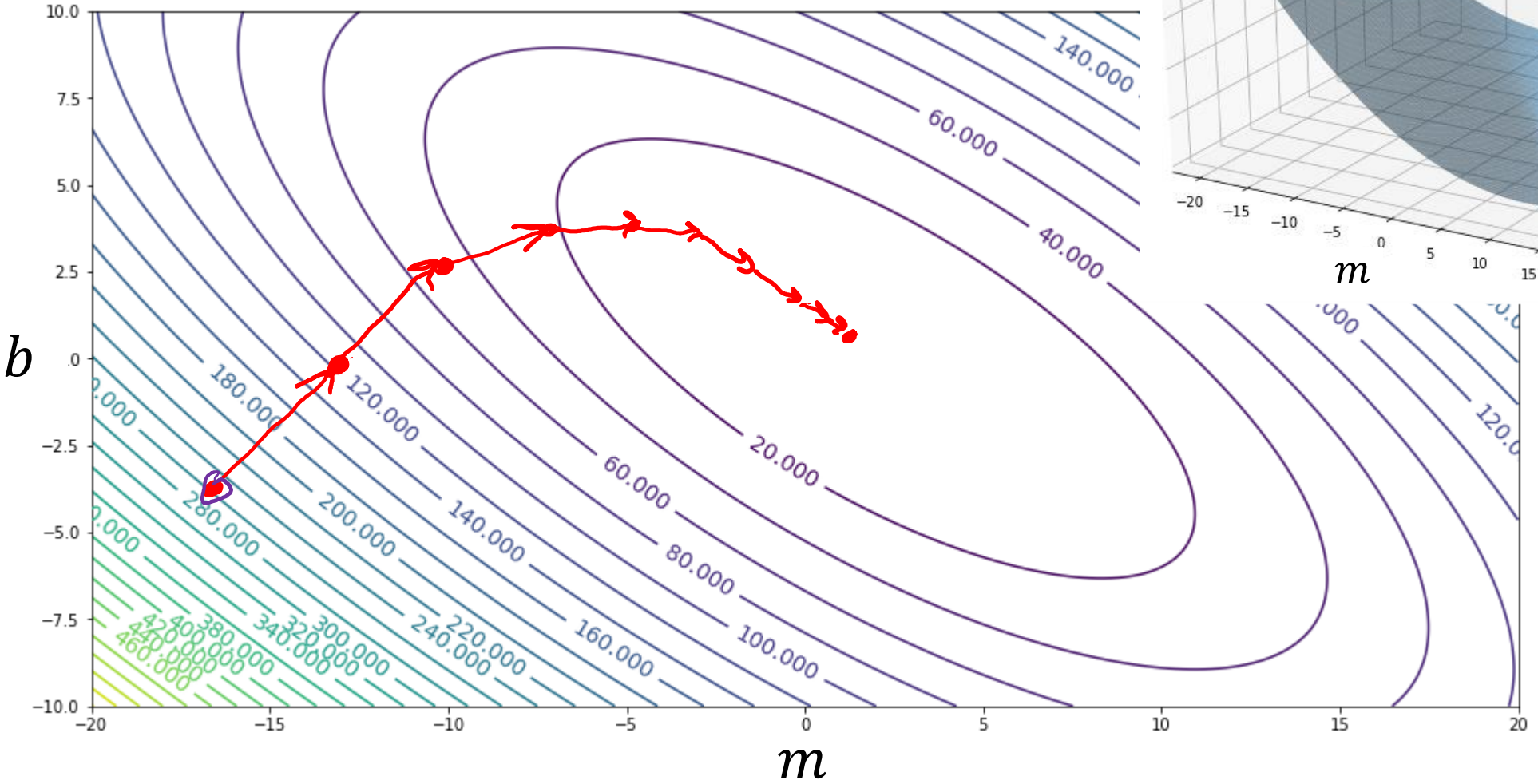
Searching for parameters that minimize mean squared error

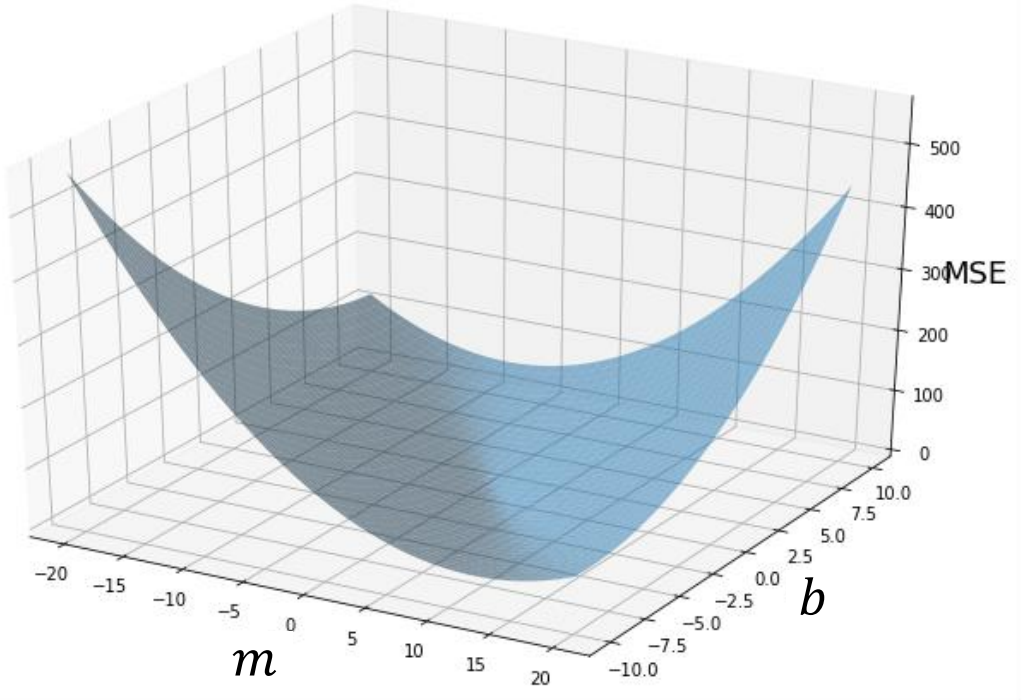




Reminder: Linear Regression

Gradient descent



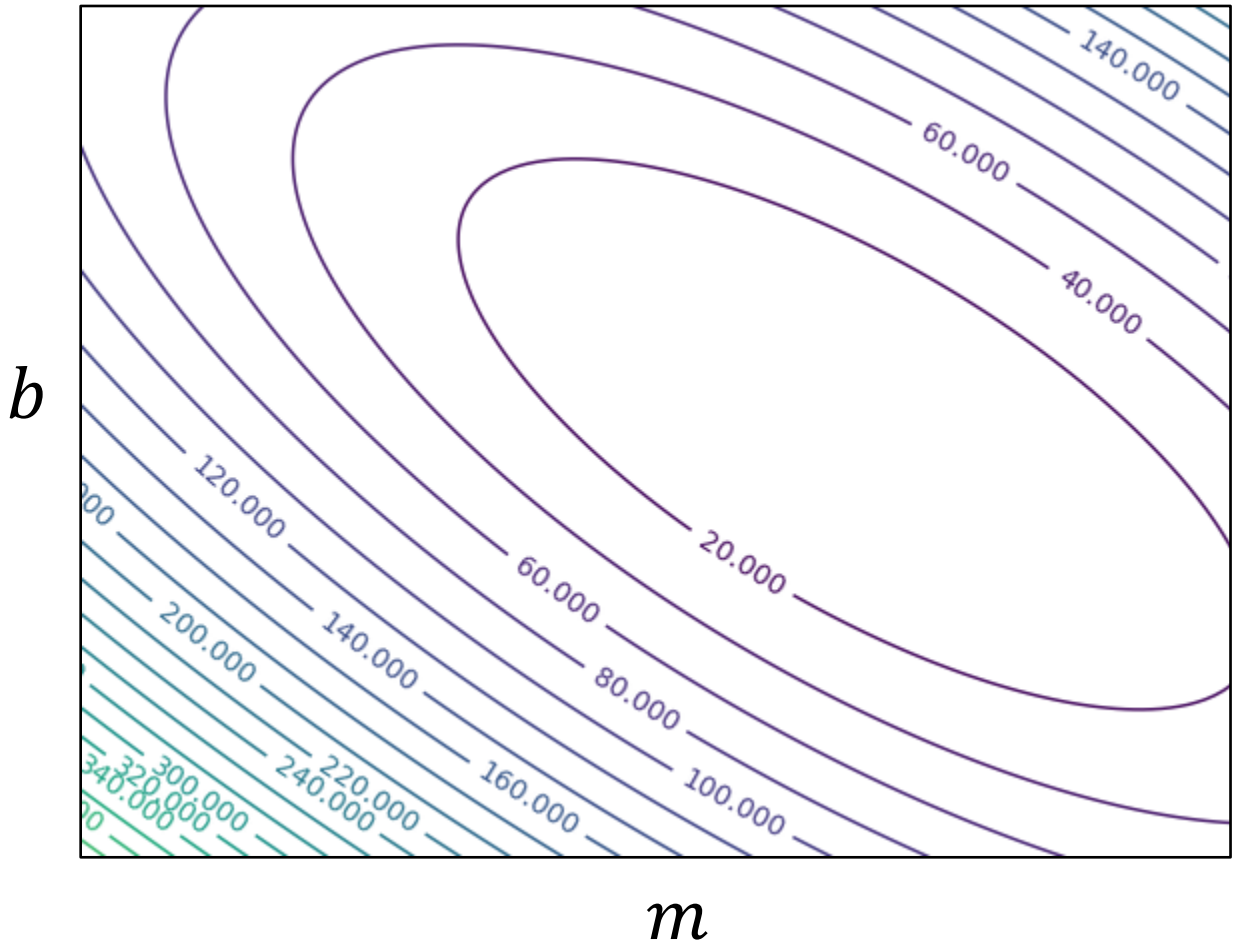


Reminder: Linear Regression

Methods for optimizing the objective

- Grid search
- Random search
- (Solve for the minimum of this paraboloid)
- Gradient descent

Performance measure:
Mean Squared Error

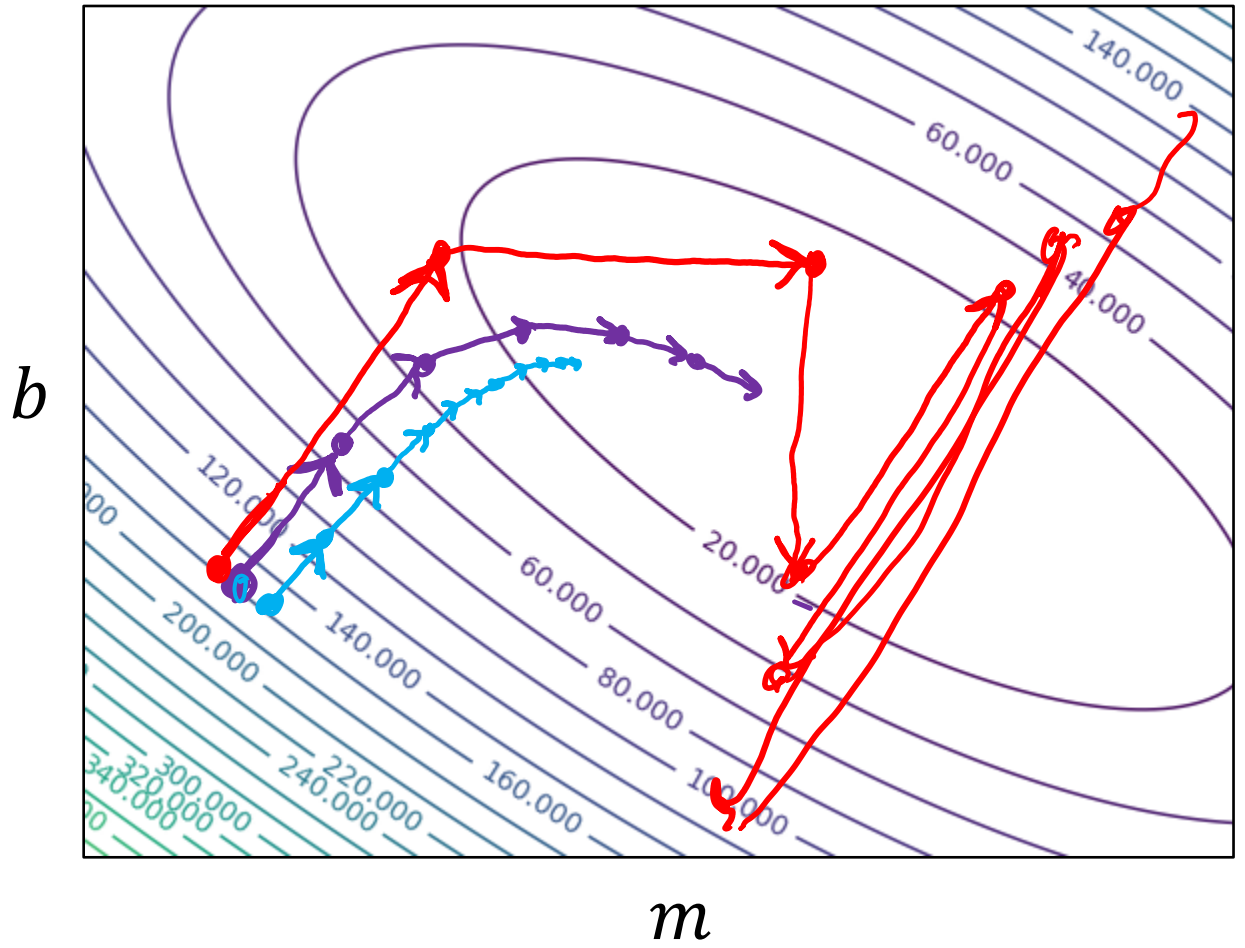


Reminder: Linear Regression

Methods for optimizing the objective

- Grid search
- Random search
- (Solve for the minimum of this parabaloid)
- Gradient descent
  - Low learning rate
  - Med learning rate
  - High learning rate

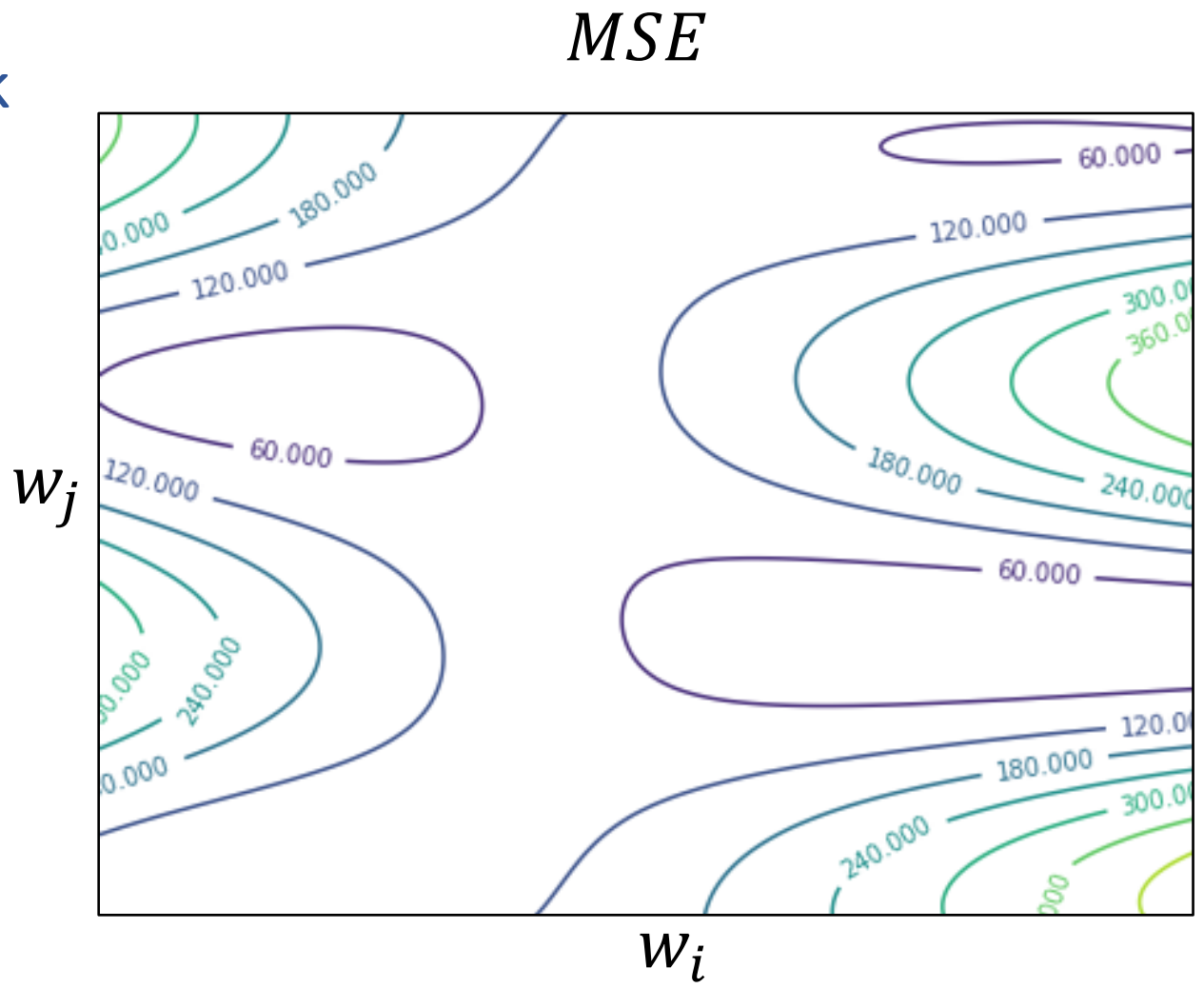
Performance measure:
Mean Squared Error



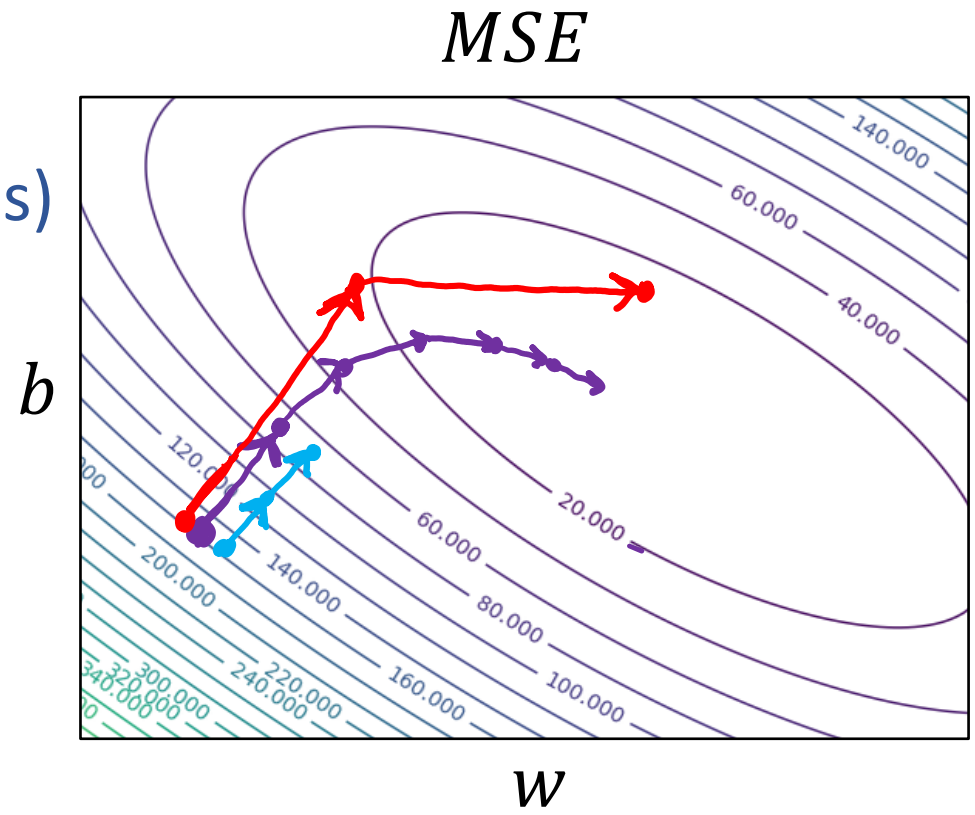
Neural Network Optimization

Methods for minimizing the mean squared error for a neural network

- Grid search
- Random search
- ☹ Can't solve for the minimum
- Gradient descent
  - Can get stuck
  - Different starting points can lead to different results



Neural Network Optimization



Gradient descent (the same, just more params)

Initialize:

- Select a learning rate, α
- $w \leftarrow random$
- $b \leftarrow random$

Repeat:

1. Calculate gradient: $grad_w, grad_b$
2. New parameter values by subtracting the gradient (scaled by the learning rate):
   - $w_{new} \leftarrow w_{prev} - \alpha \cdot grad_w$
   - $b_{new} \leftarrow b_{prev} - \alpha \cdot grad_b$