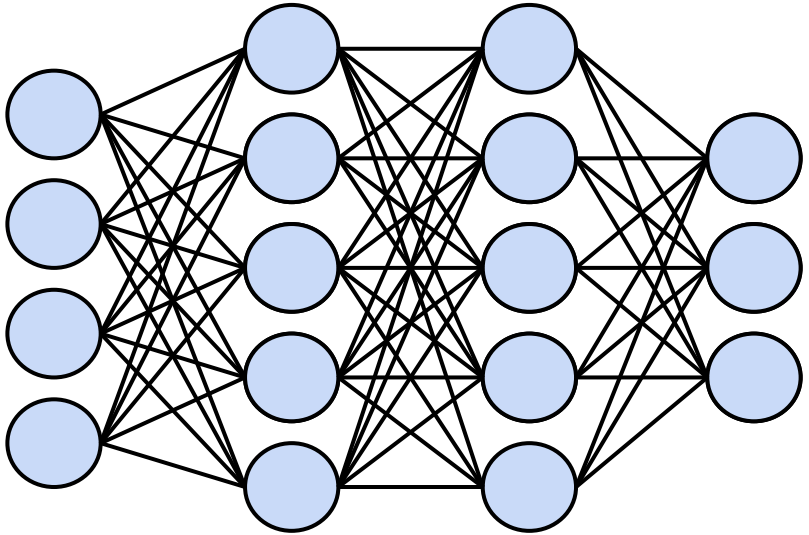
Introduction to Neural Networks

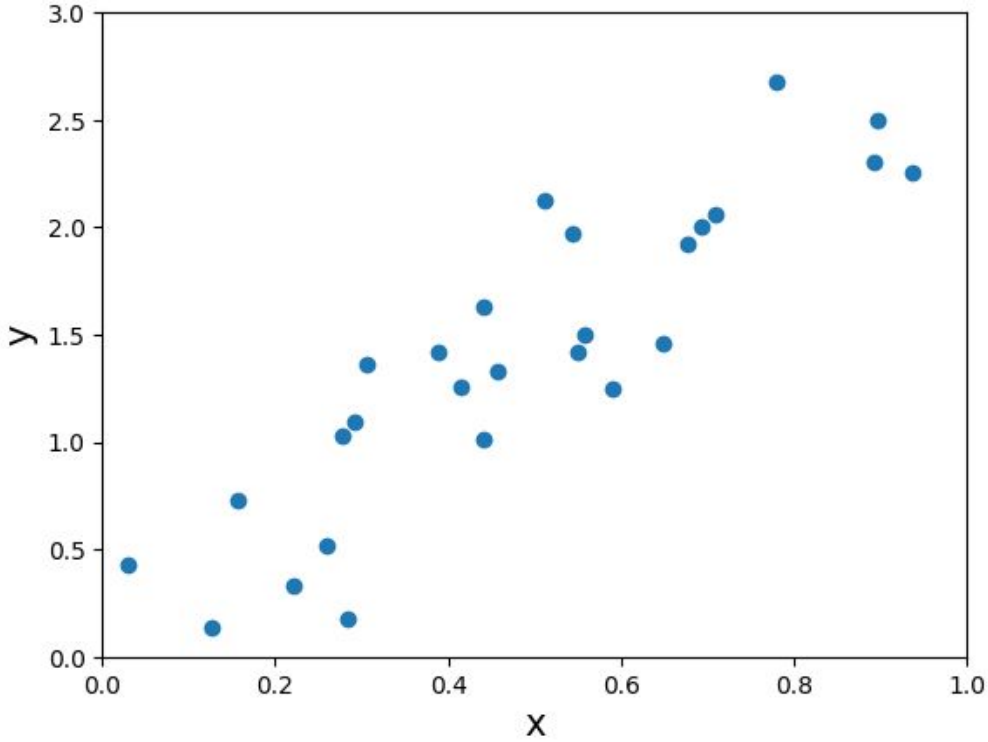
David, Michael, Teddy, Tommy

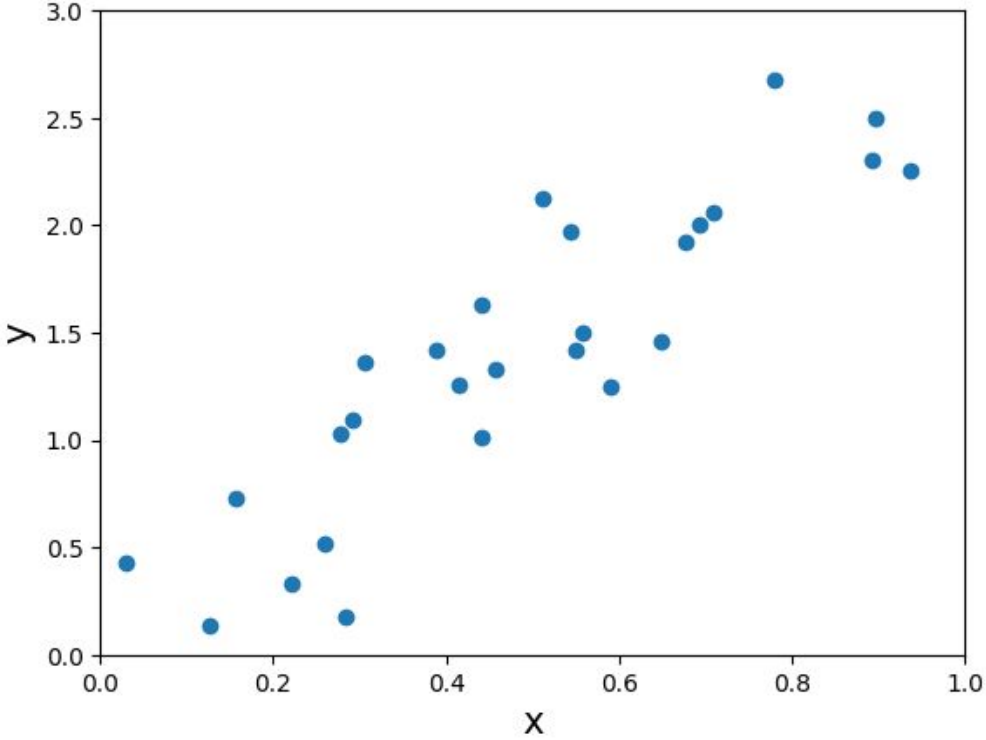


What is a Neural Network

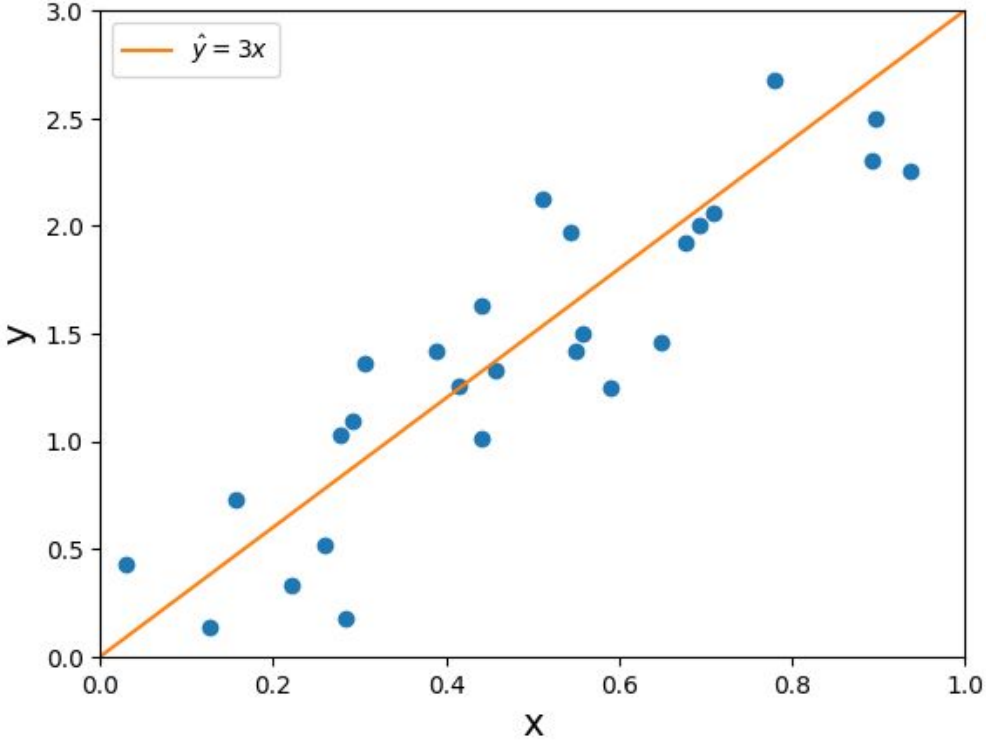
- Machine Learning
  - Regression
  - Classification
- Applications
  - Image Classification
  - Speech Recognition
  - Large Language Models

Neural Network Architecture

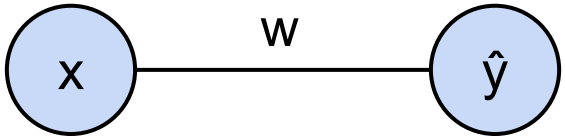

3.0
2.5
2.0
1.5
1.0
0.5
0.0
y
0.0
0.2
0.4
0.6
0.8
1.0
x



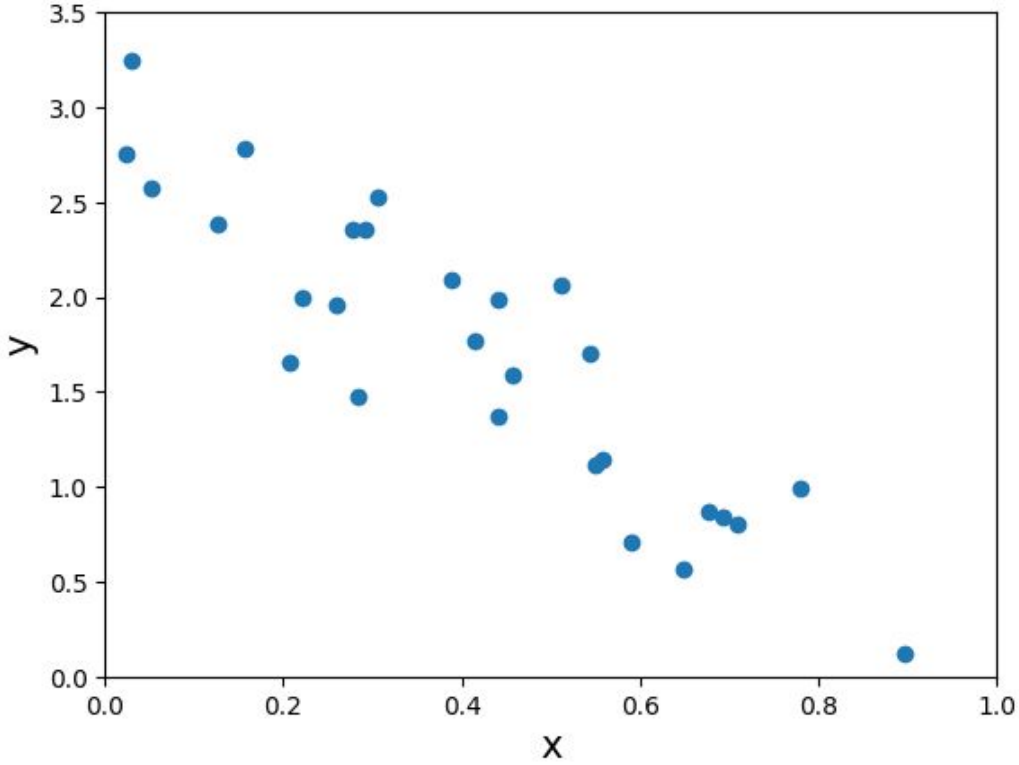
$$\hat{y} = wx$$

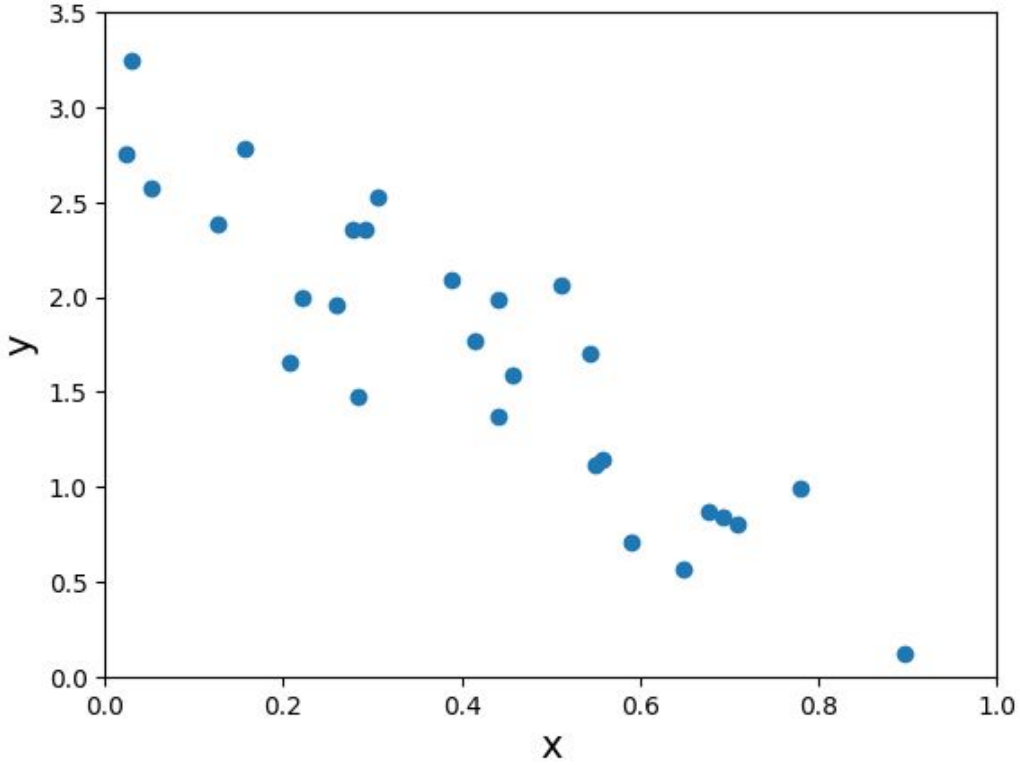


$$\hat{y} = wx$$

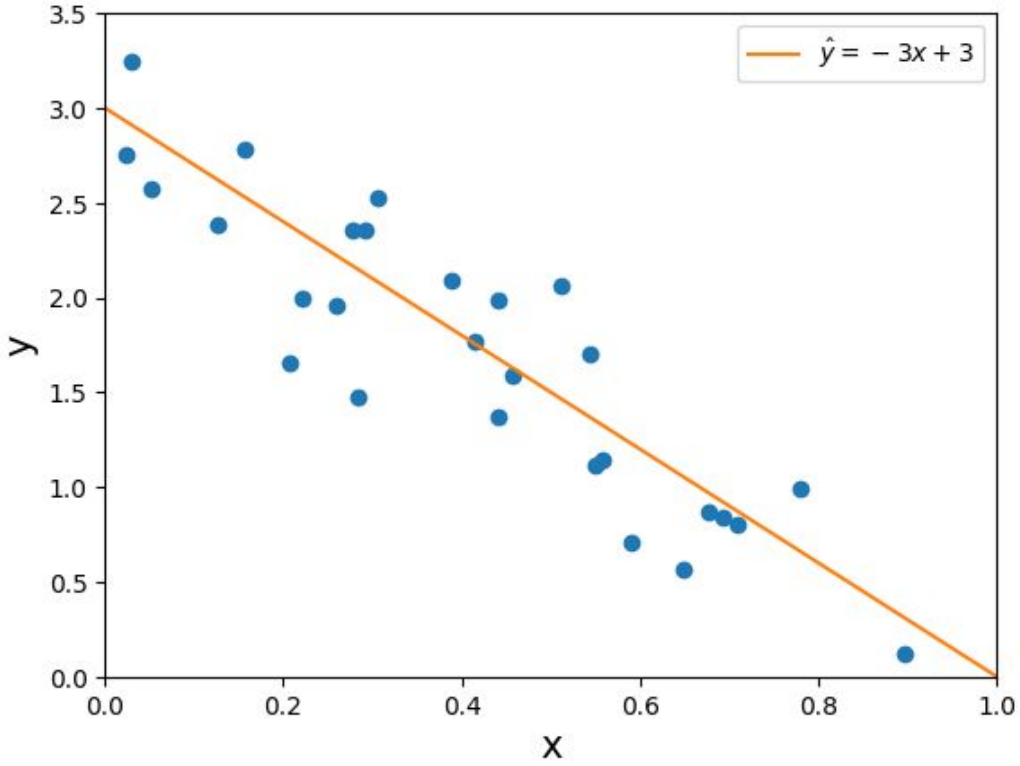


$$\hat{y} = wx$$

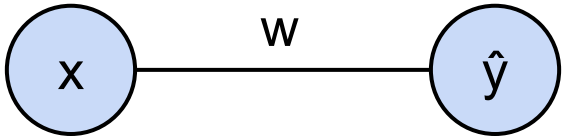

3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
y
0.0
0.2
0.4
0.6
0.8
1.0
x



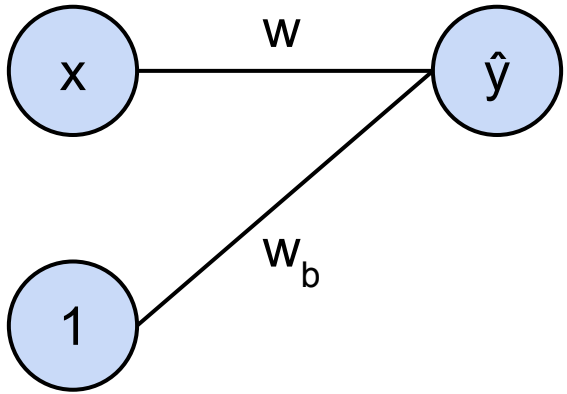
$$\hat{y} = wx + b$$



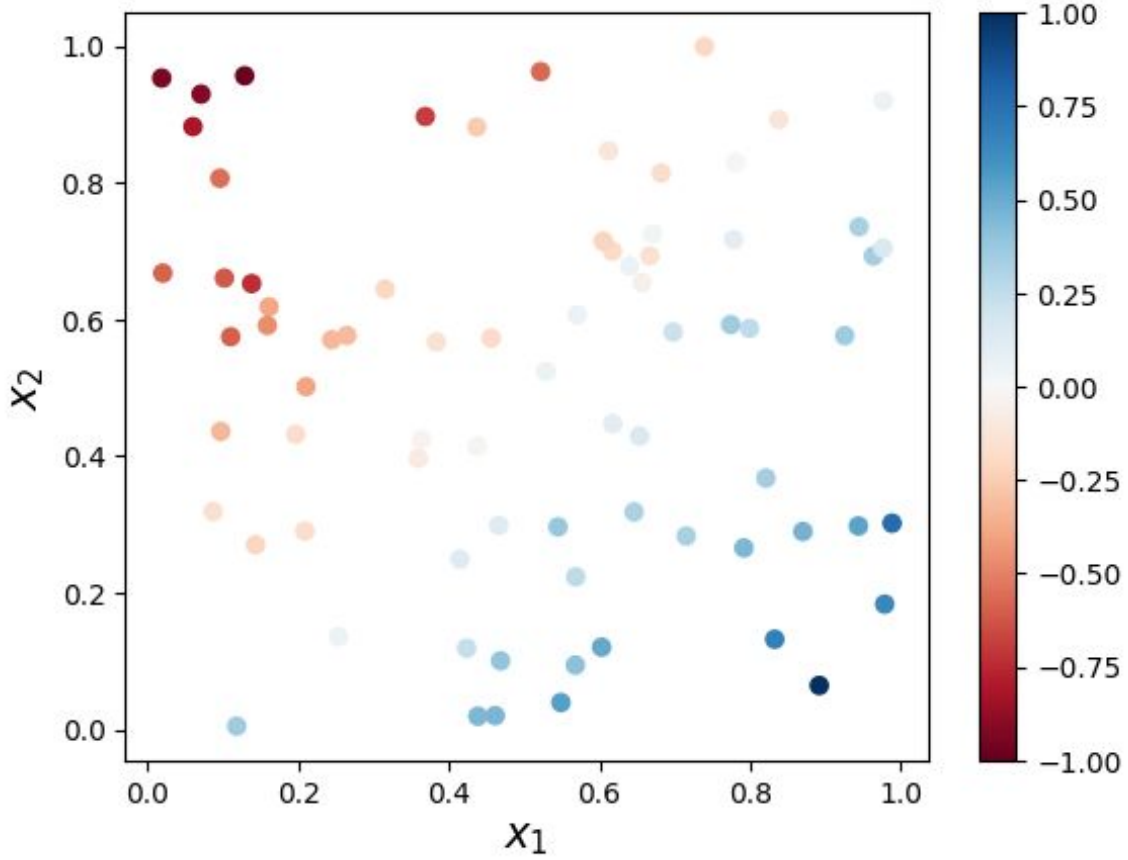
$$\hat{y} = wx + b$$

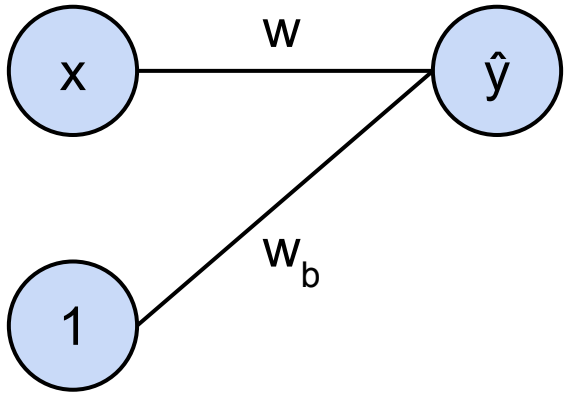


$$\hat{y} = wx$$

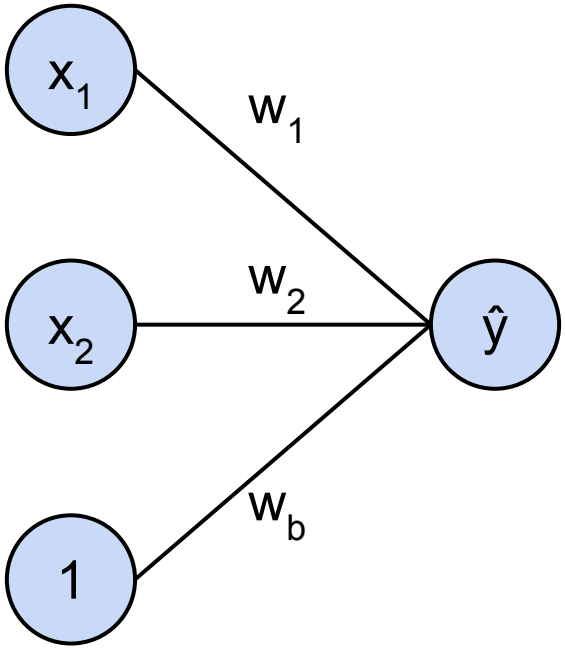


$$\hat{y} = wx + w_b$$

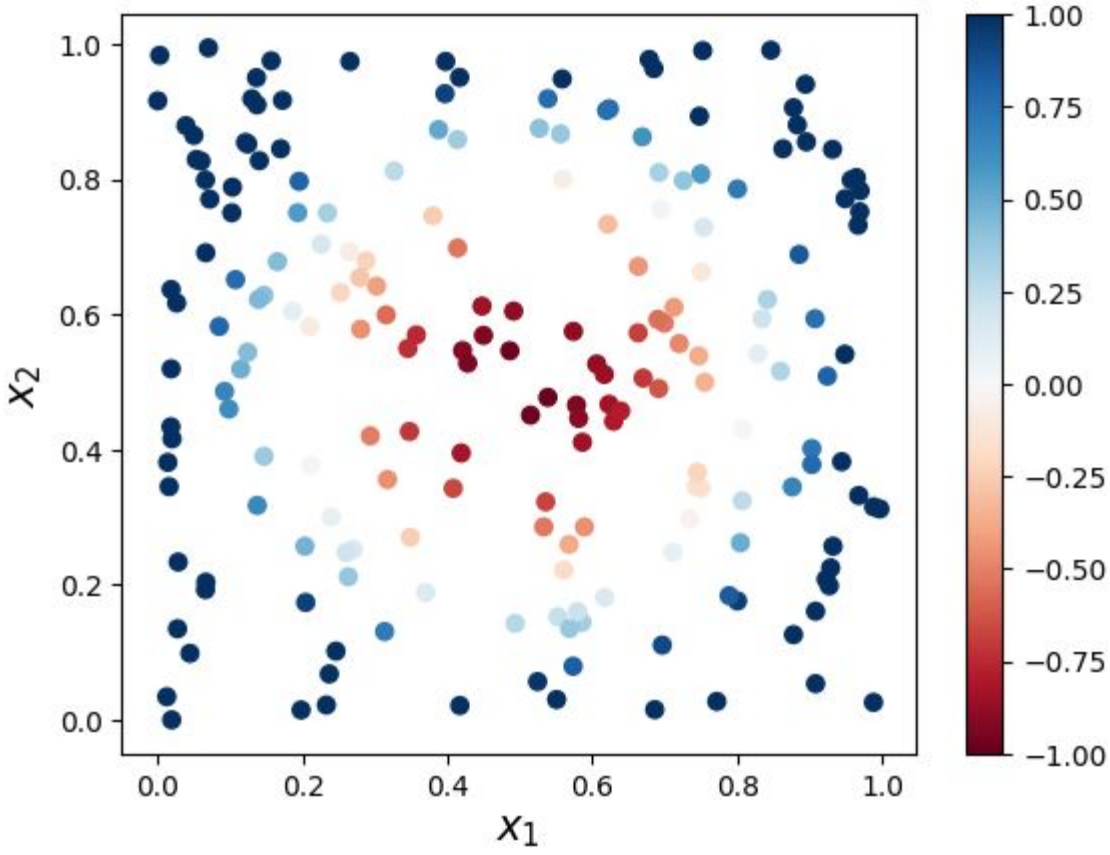

1.0
0.8
0.6
0.4
0.2
0.0
x_2
0.0
0.2
0.4
0.6
0.8
1.0
x_1
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00

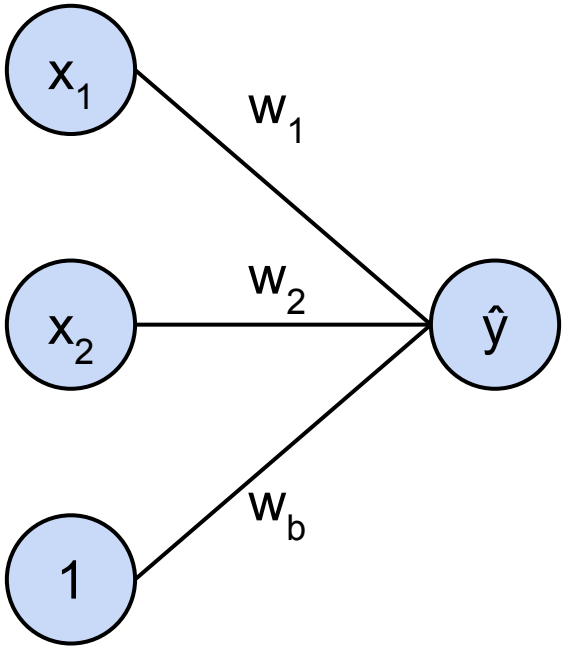


$$\hat{y} = wx + w_b$$

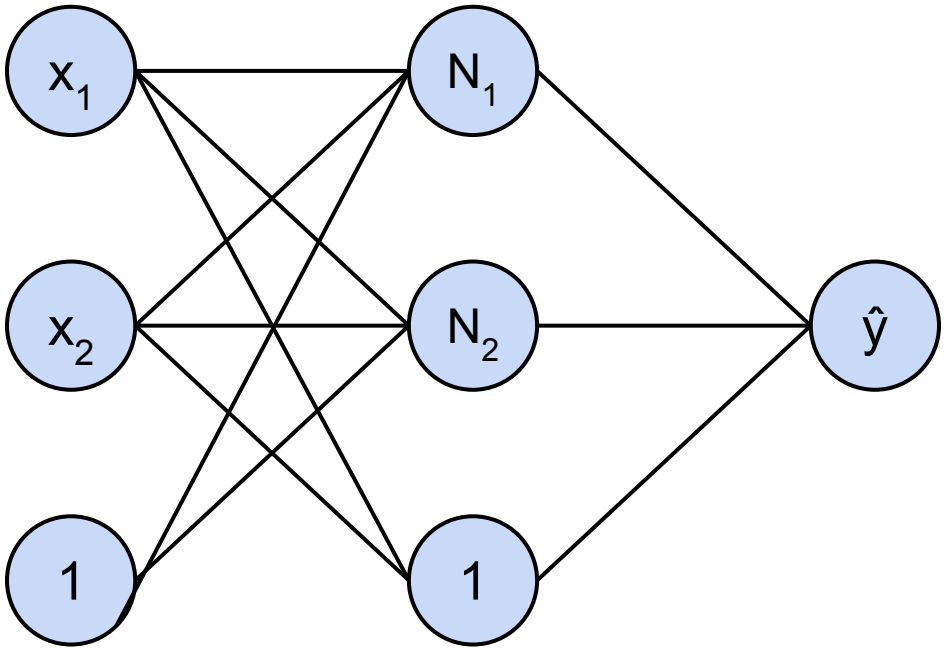


$$\hat{y} = w_1 x_1 + w_2 x_2 + w_b$$

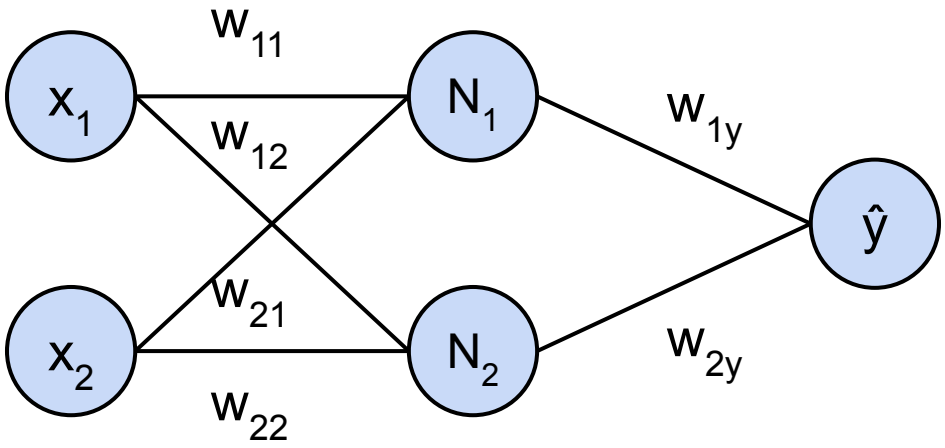

1.0
0.8
0.6
0.4
0.2
0.0
x_2
0.0
0.2
0.4
0.6
0.8
1.0
x_1
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00



$$\hat{y} = w_1 x_1 + w_2 x_2 + w_b$$



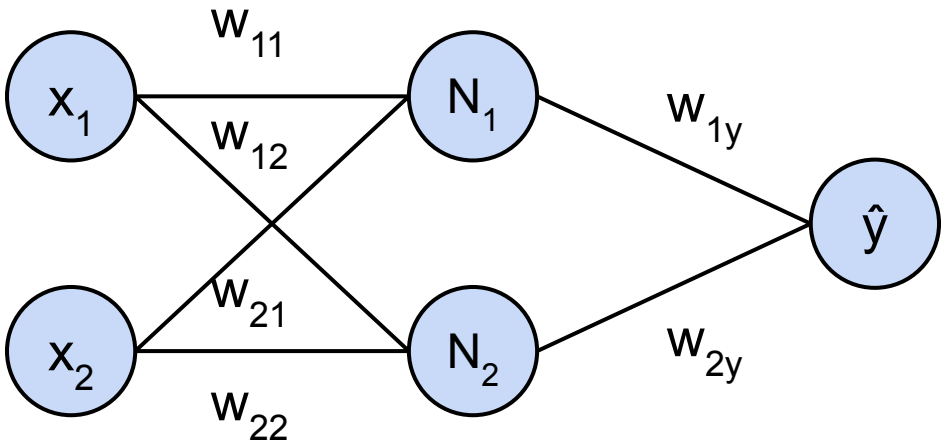
Hidden Layer



$$N_1 = w_{11}x_1 + w_{21}x_2 + b_{N1}$$

$$N_2 = w_{12}x_1 + w_{22}x_2 + b_{N2}$$

$$\hat{y} = w_{1y}N_1 + w_{2y}N_2 + b_y$$

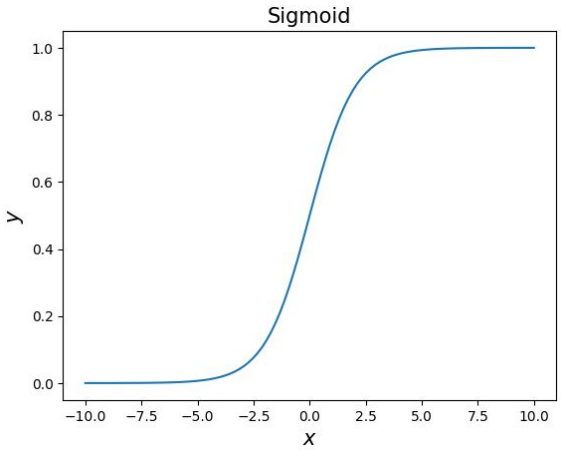


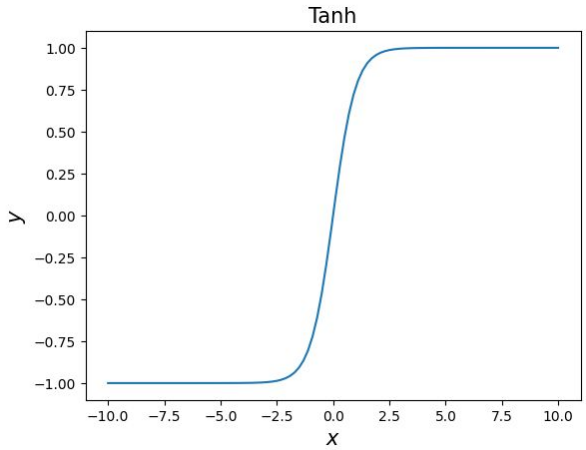
$$N_1 = \sigma(w_{11}x_1 + w_{21}x_2 + b_{N1})$$

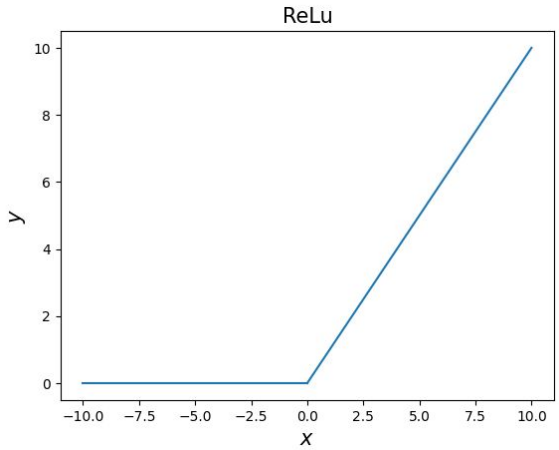
$$N_2 = \sigma(w_{12}x_1 + w_{22}x_2 + b_{N2})$$

$$\hat{y} = \sigma(w_{1y}N_1 + w_{2y}N_2 + b_y)$$

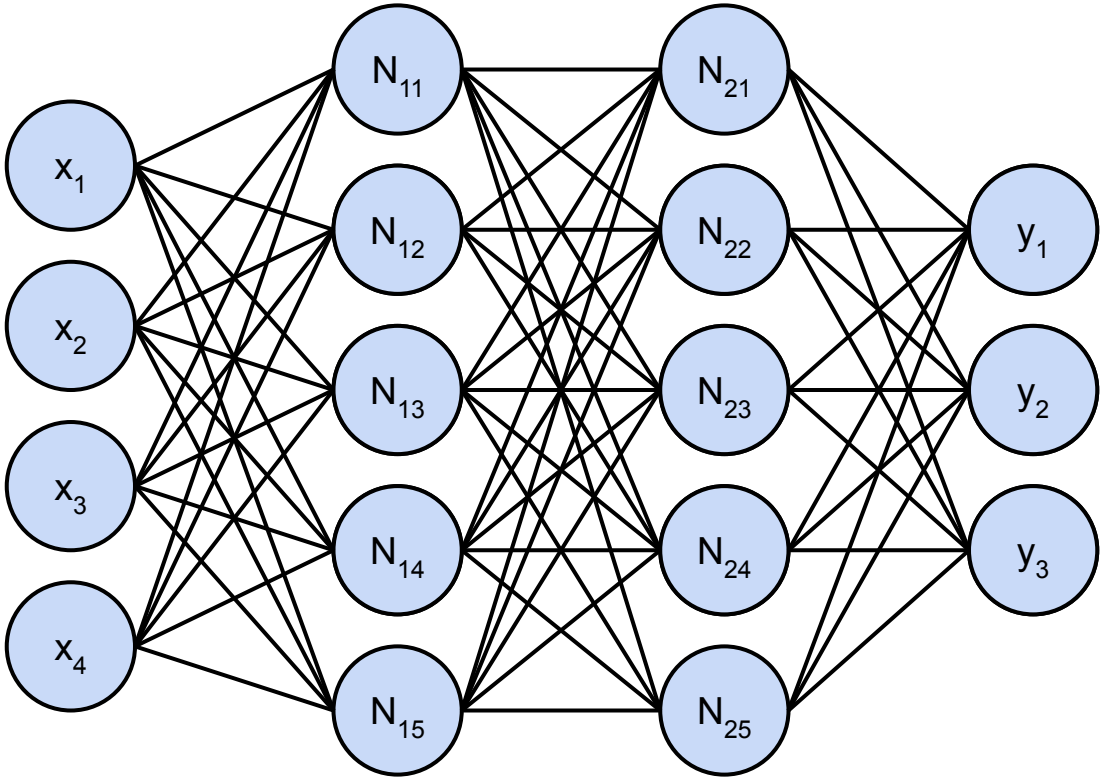
Activation Functions







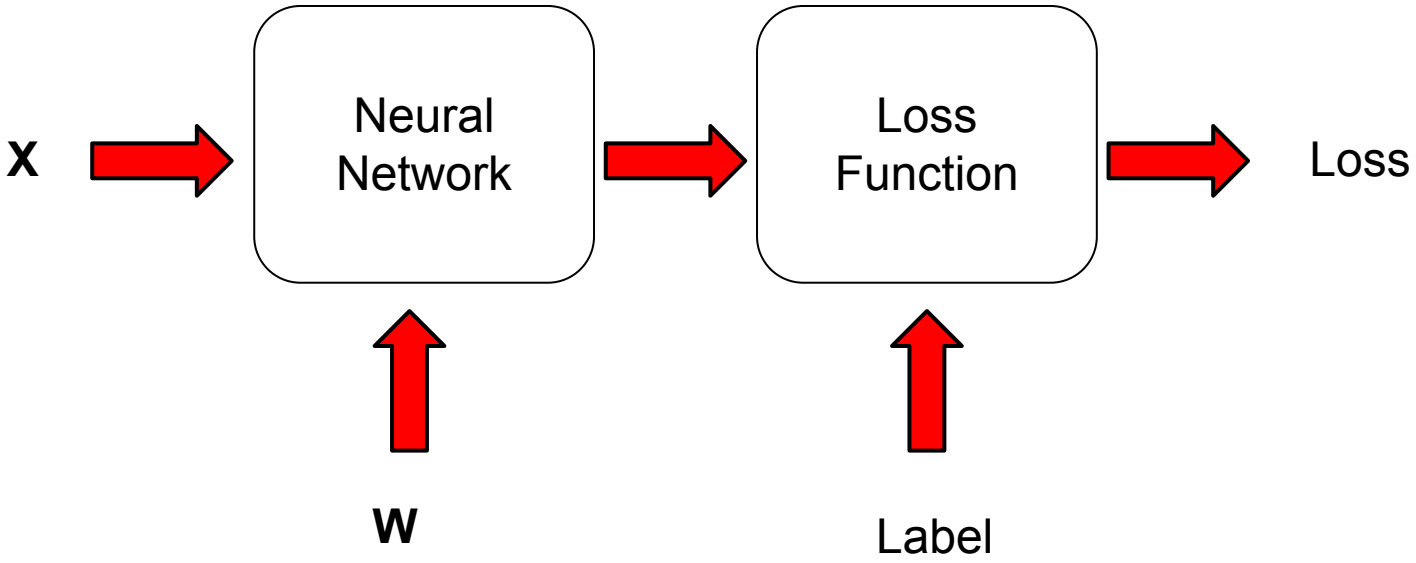
See https://keras.io/api/layers/activations/ for more common activation functions


x1
x2
x3
x4
N11
N12
N13
N14
N15
N21
N22
N23
N24
N25
y1
y2
y3

Training Neural Networks

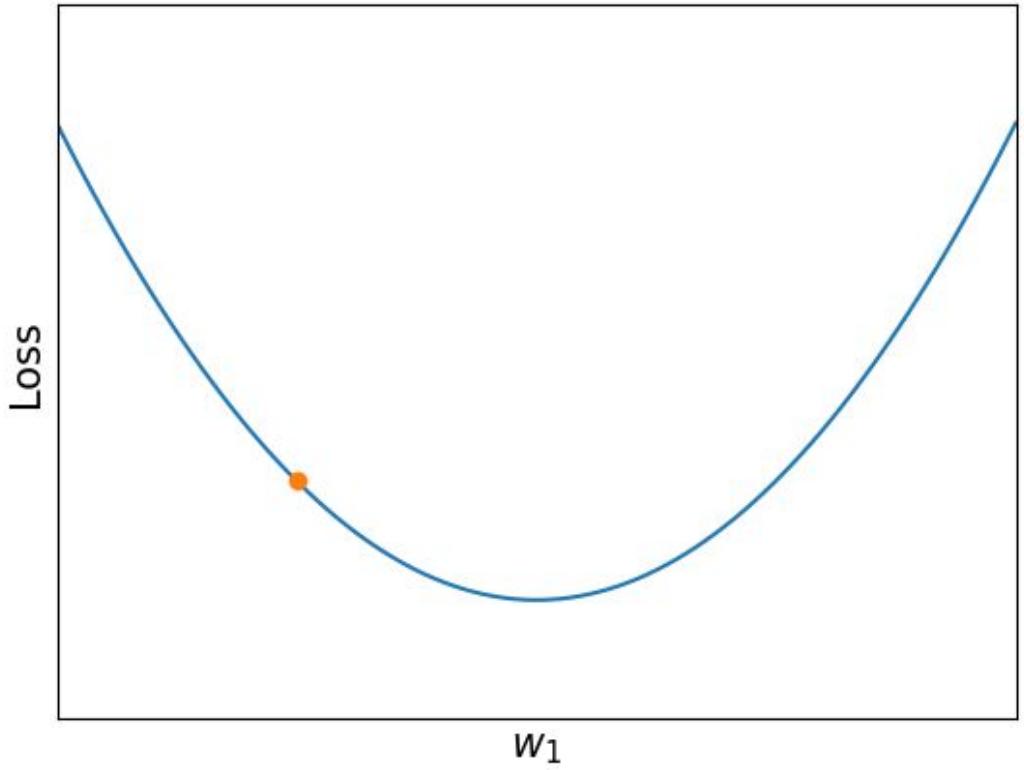
Loss Functions

- Regression
  - Mean Squared Error
  - Mean Absolute Error
- Classification
  - Cross entropy
- See https://keras.io/api/losses/ for more examples

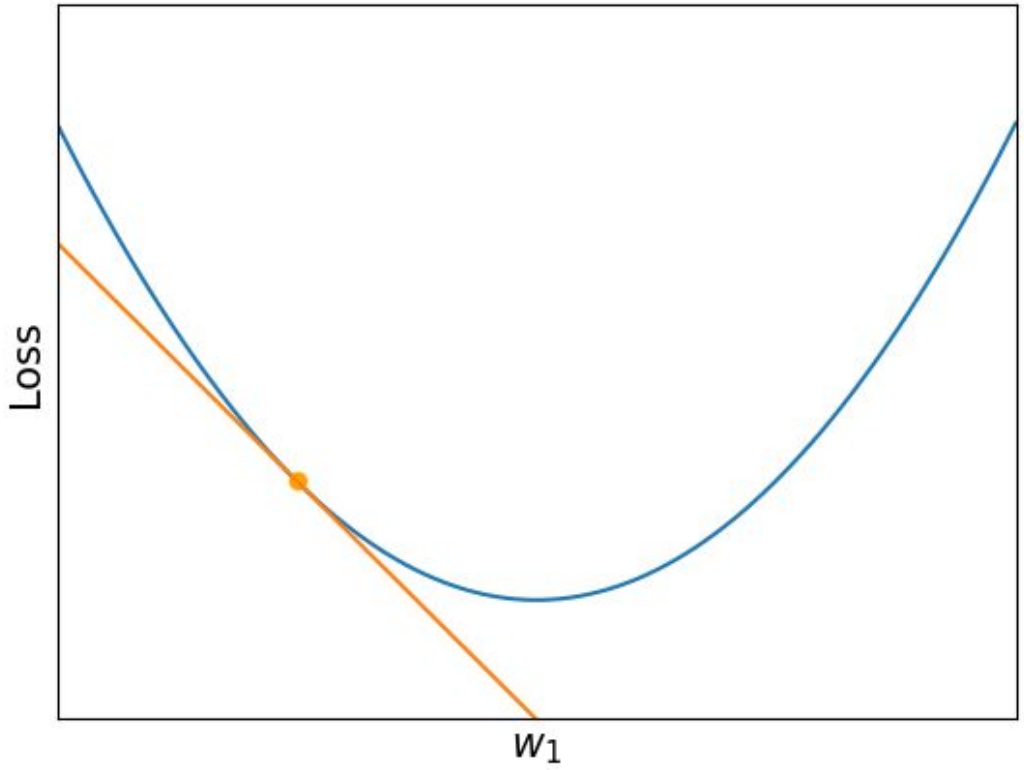


f(**W**): Loss as a function of weights for a given **X**

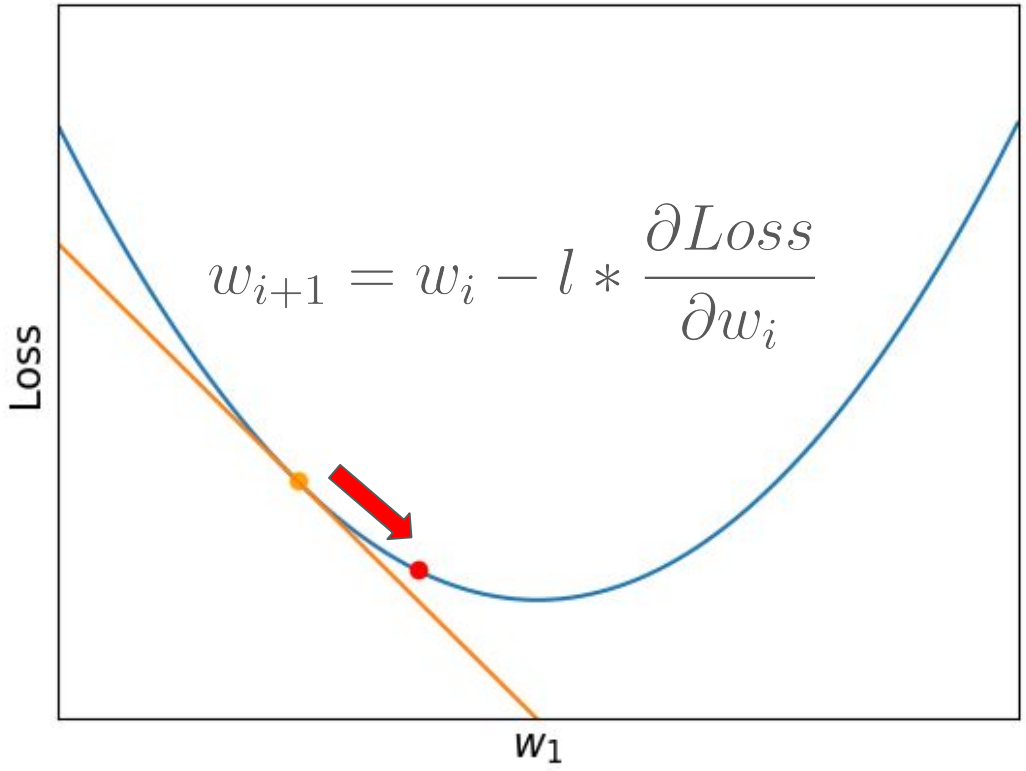
Gradient Descent



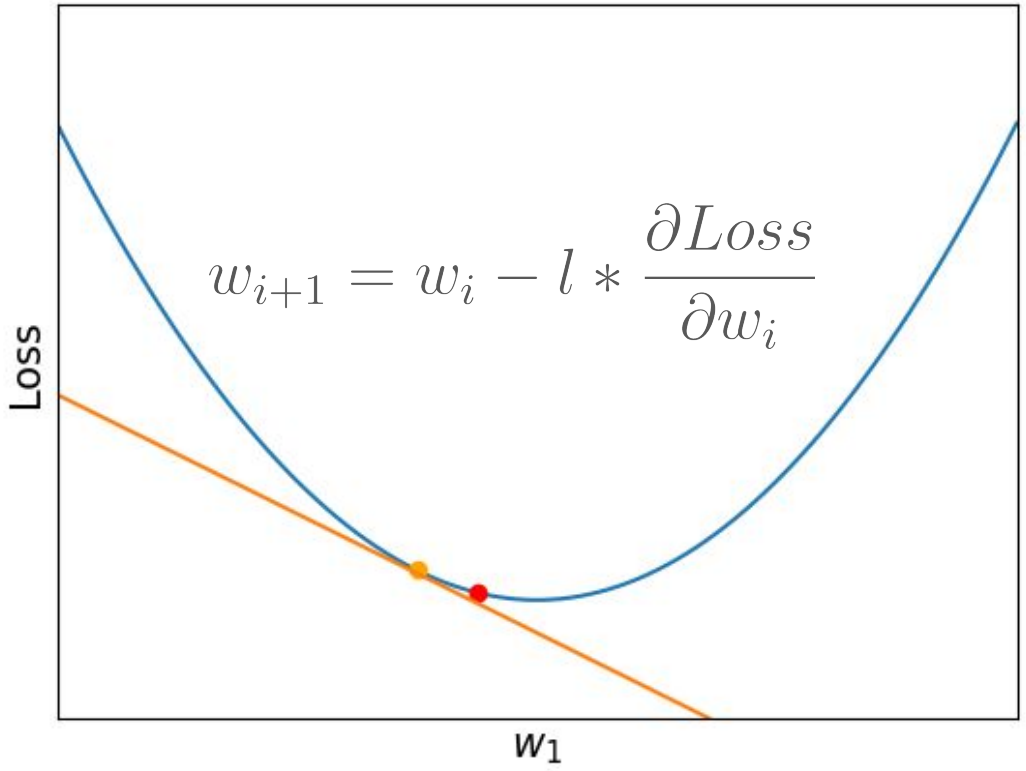
Gradient Descent



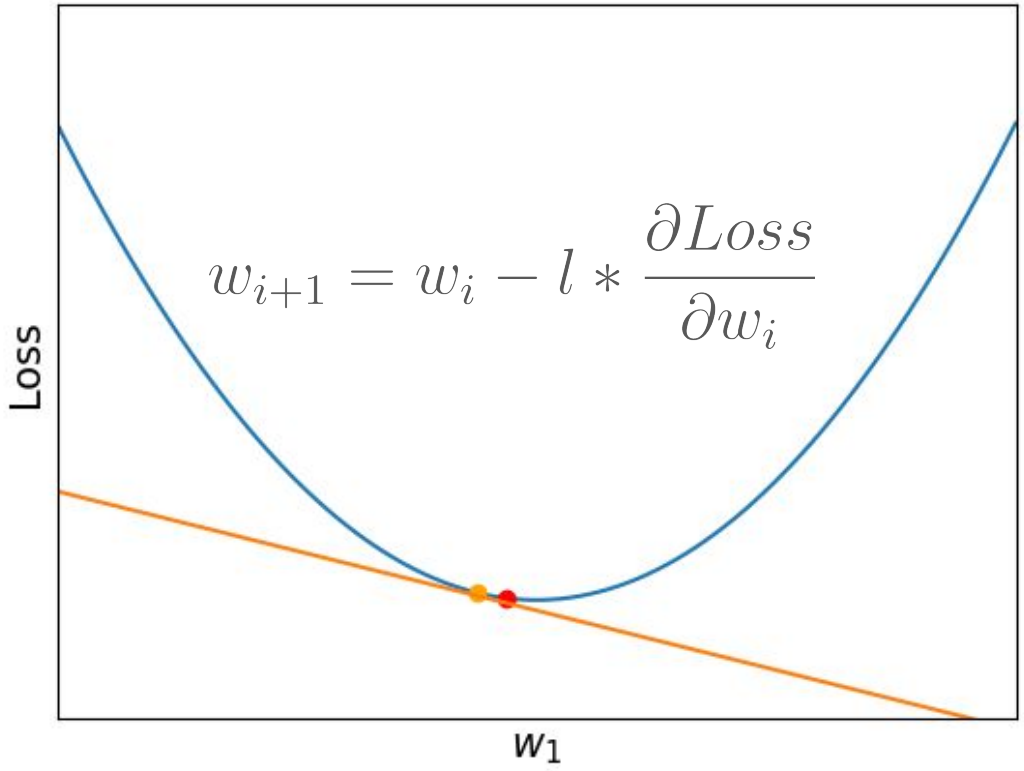
Gradient Descent



Gradient Descent



Gradient Descent



Optimizers

- Stochastic Gradient Descent (SGD)
- Adam
- RMSProp
- See https://keras.io/api/optimizers/ for other common optimizers