Machine Learning
Live Session #4

Review of
Training Neural Networks

Neural Networks and Deep Learning

- Weights W are initialized randomly
  - Unless another initialization is specified (e.g., from a previous model or experience)
- Then, compare the predictions of the model with the actual labels in the training data
- Based on the comparison, update the weights to get the predictions closer to the actual labels
- Repeat the previous two steps until stopping criterion is met
- How do we make the comparison and update the weights to find the optimal values W^*?

Neural Networks and Deep Learning

- How do we make the comparison?
  - This is the role of the loss function, which is *only* a function of the weights W, given the training data and architecture of the NN

$$Loss(W; training\ set, architecture\ of\ NN)$$

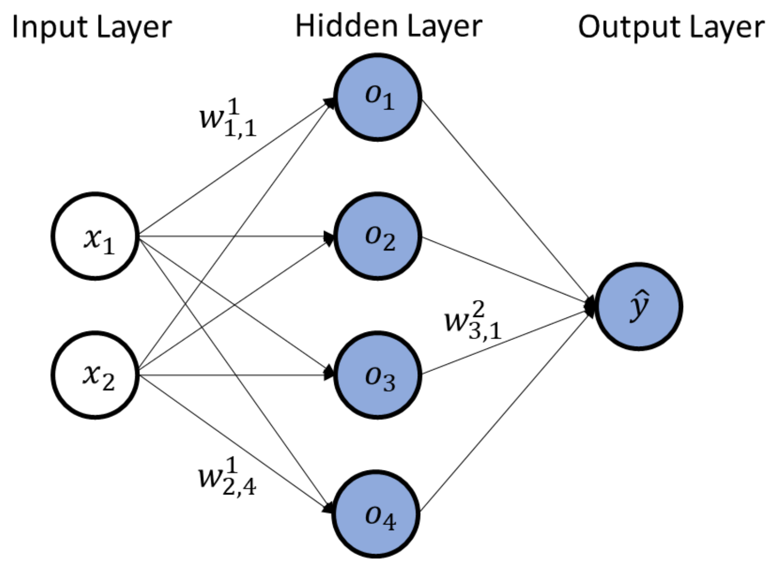
  - Larger value of the loss means the compared values were further away
  - Smaller value of the loss means the compared values were closer

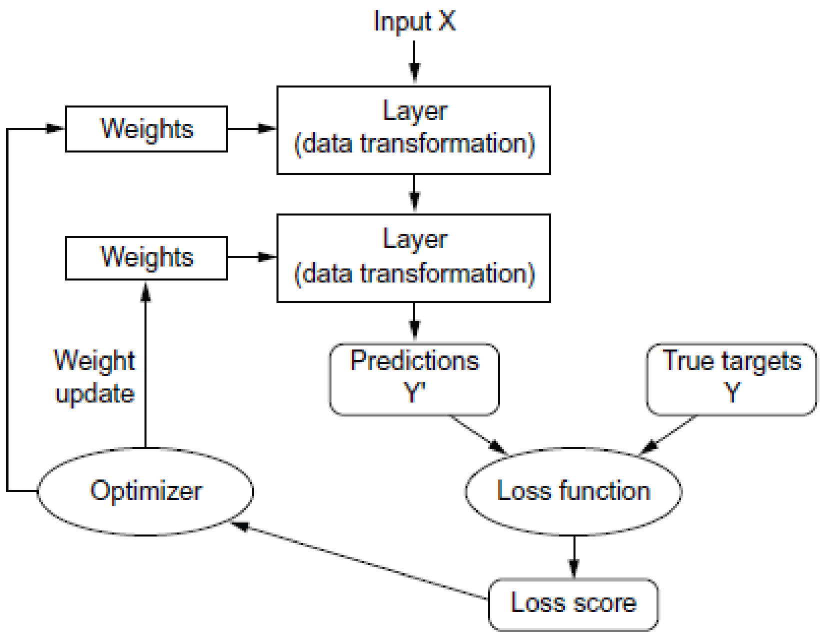
Details of Training Process

- Important: for a particular training set and architecture, the loss is only a function of the weights W

$$Loss(W; training\ set, architecture\ of\ NN)$$

- The loss function is used along with an optimization algorithm to determine how to change the weights to improve the loss
- The goal of the optimization is to find the optimal weight values W^* that minimize the loss





Neural Networks and Deep Learning

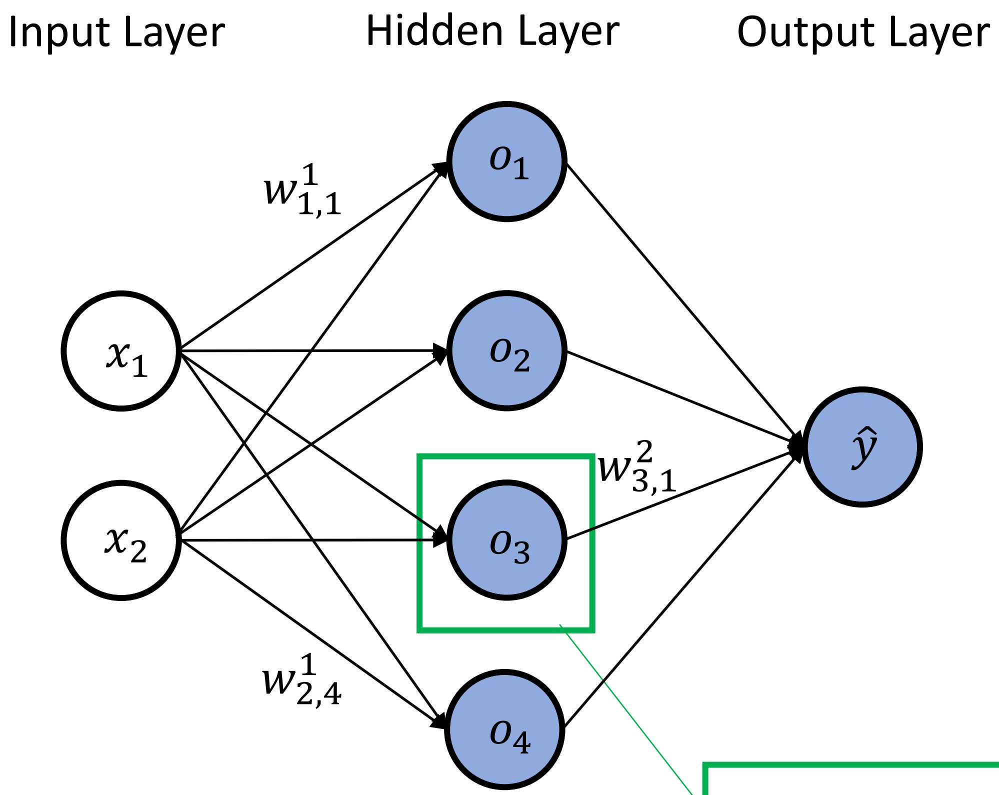
- To calculate the loss for a given set of weight values W
  1. Each x in the training set is put through the network to obtain a prediction $\hat{y}(x\,; W)$
  2. Calculate the loss function (example below using mean squared loss function for regression)

$$Loss(W; training\ set, architecture\ of\ NN) = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}(x_i\,; W)\ - y_i)^2$$

$x_1, x_2, \ldots, x_n$ are training observations and $y_1, y_2, \ldots, y_n$ are actual labels

Training Dense Feed-Forward Neural Networks

Income (x_1)	Age (x_2)
22003	45
57230	54
75137	28
31208	54
54078	23
44413	44
55237	46



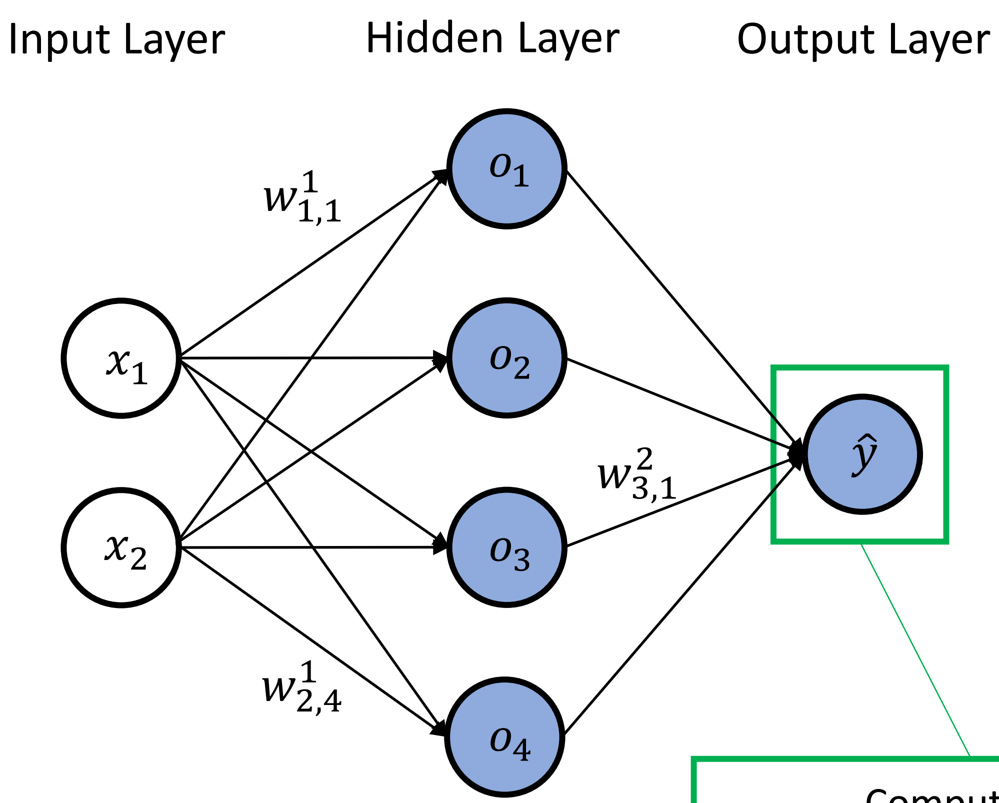
Computation Node

$$o_3 = f_1(Weights \times Inputs + Bias)$$

$$o_3 = f_1(w_{1,3}^{1} \times x_1 + w_{2,3}^{1} \times x_2 + w_{0,3}^{1})$$

Training Dense Feed-Forward Neural Networks

Income (x_1)	Age (x_2)
22003	45
57230	54
75137	28
31208	54
54078	23
44413	44
55237	46



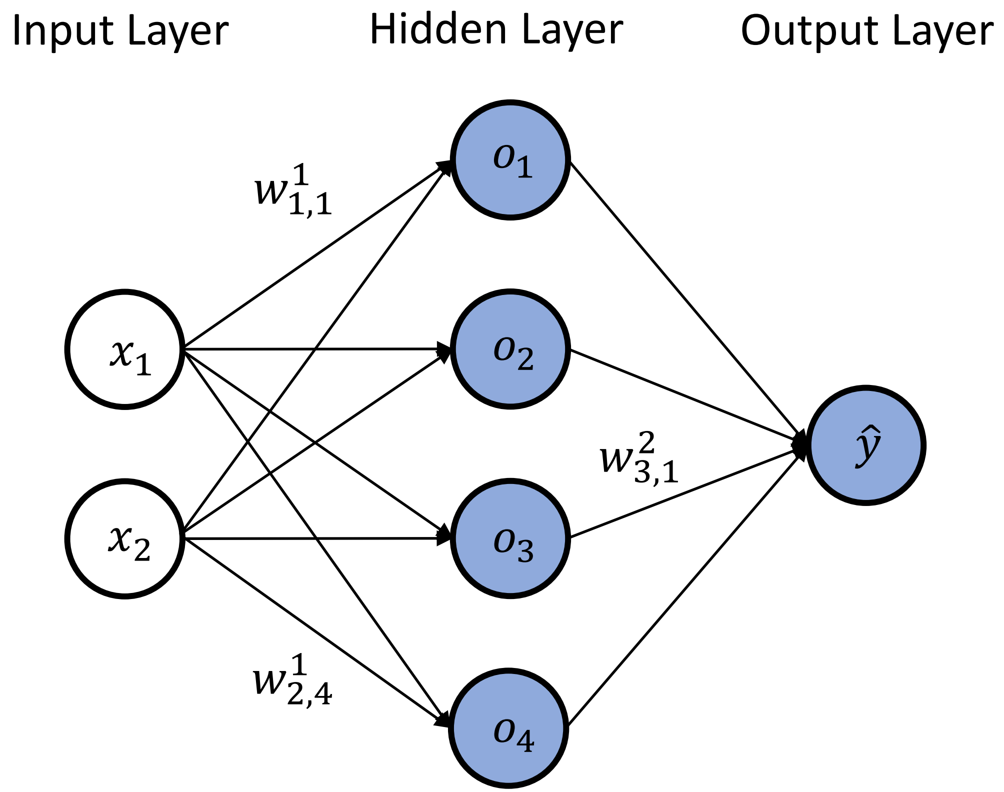
Computation Node

$$\hat{y} = f_2(Weights * Inputs + Bias)$$

$$\hat{y} = f_2(w^2_{1,1}o_1 + w^2_{2,1}o_2 + w^2_{3,1}o_3 + w^2_{4,1}o_4 + w^2_0)$$

Training Dense Feed-Forward Neural Networks

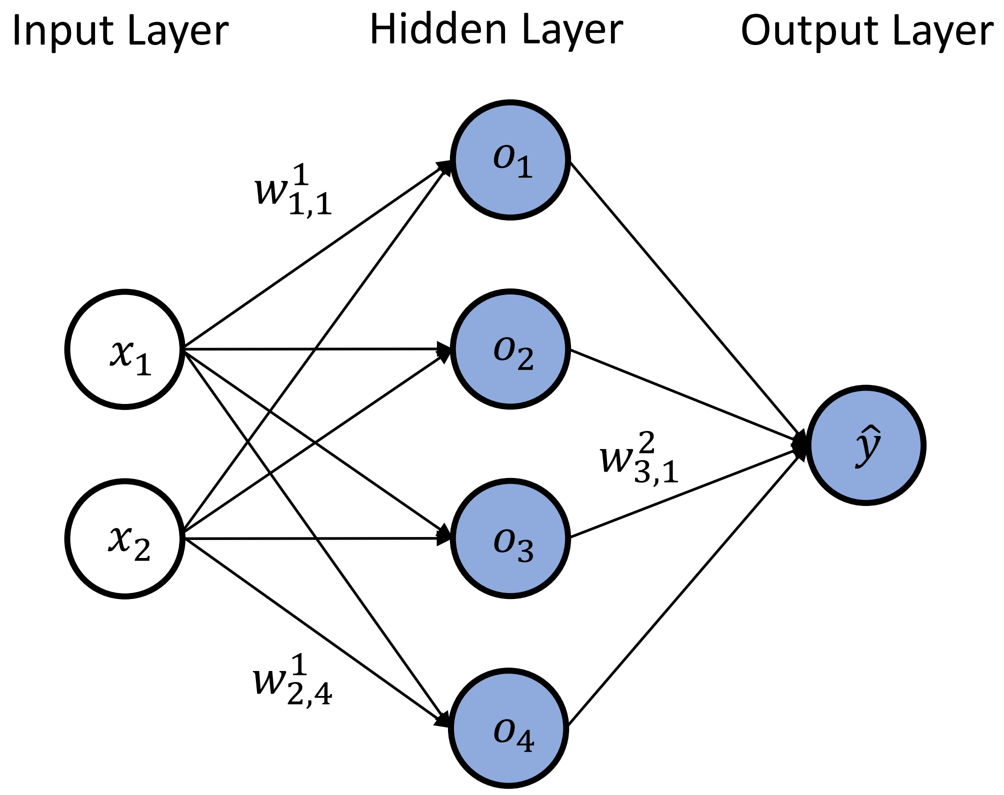
Income (x_1)	Age (x_2)
22003	45
57230	54
75137	28
31208	54
54078	23
44413	44
55237	46



Predicted Revenue Given Set of Weights W
$\hat{y}(22003{,}45; W)$
$\hat{y}(57230{,}54; W)$
$\hat{y}(75137{,}28; W)$
$\hat{y}(31208{,}54; W)$
$\hat{y}(54078{,}23; W)$
$\hat{y}(44413{,}44; W)$
$\hat{y}(55237{,}46; W)$

Training Dense Feed-Forward Neural Networks

Income (x_1)	Age (x_2)
22003	45
57230	54
75137	28
31208	54
54078	23
44413	44
55237	46



Update weights based on loss function and optimizer

Predicted Revenue Given Set of Weights W
$\hat{y}(22003{,}45; W)$
$\hat{y}(57230{,}54; W)$
$\hat{y}(75137{,}28; W)$
$\hat{y}(31208{,}54; W)$
$\hat{y}(54078{,}23; W)$
$\hat{y}(44413{,}44; W)$
$\hat{y}(55237{,}46; W)$

The loss function compares these

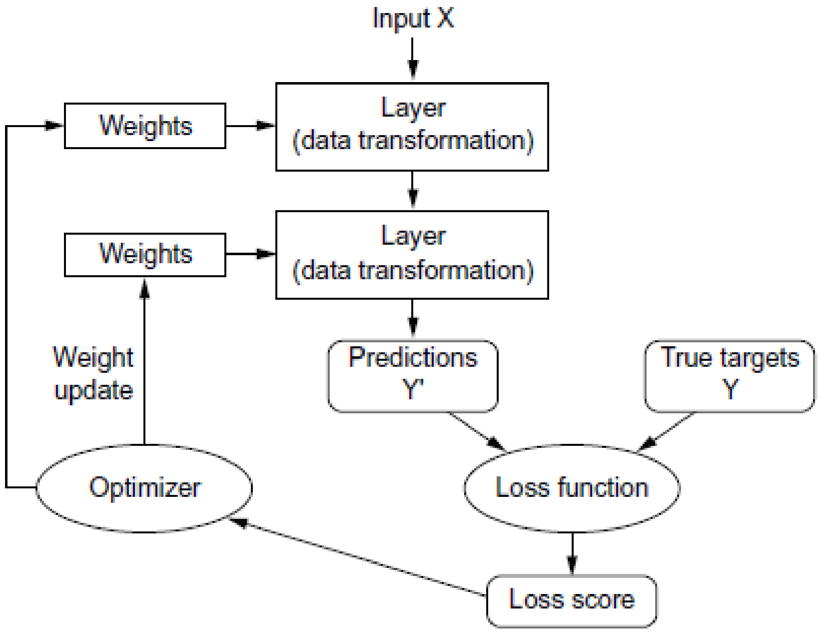
Revenue from Customer
14.03875
23.31168
24.05046
18.5386
18.50195
20.63106
22.32953

$$Loss(W; training\ set, architecture\ of\ NN) = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}(x_i\,; W)\ - y_i)^2$$

Neural Networks and Deep Learning

- When the training set is too large, it's not feasible to calculate the loss based on the entire training set every iteration
  - Solution: use batches consisting of subsets of rows of training data
  - Take for example a batch size of 2

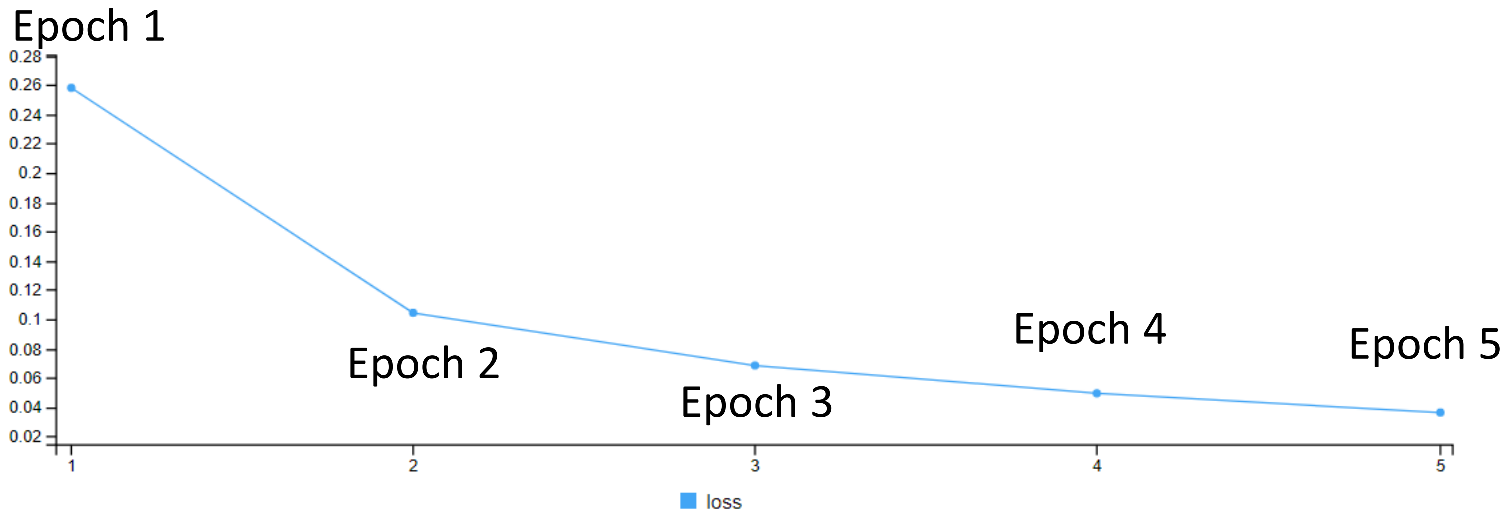
Income (Customer Feature 1)	Age (Customer Feature 2)	Revenue from Customer	Predicted Revenue Given Set of Weights W
22003	45	14.03875	$\hat{y}(22003{,}45; W)$
57230	54	23.31168	$\hat{y}(57230{,}54; W)$
75137	28	24.05046	$\hat{y}(75137{,}28; W)$
31208	54	18.5386	$\hat{y}(31208{,}54; W)$
54078	23	18.50195	$\hat{y}(54078{,}23; W)$
44413	44	20.63106	$\hat{y}(44413{,}44; W)$
55237	46	22.32953	$\hat{y}(55237{,}46; W)$



Neural Networks and Deep Learning

- Take for example a batch size of 2
  - Iterate through the training data taking two rows at a time
    - On each iteration, define the loss function $Loss(W | selected\ two\ rows\ of\ training\ data, architecture\ of\ NN)$
    - Use this loss function in conjunction with optimization algorithm to determine how to update weights
  - After iterating through the entire training set, we have completed an "epoch"
    - We specify the number of epochs in advance

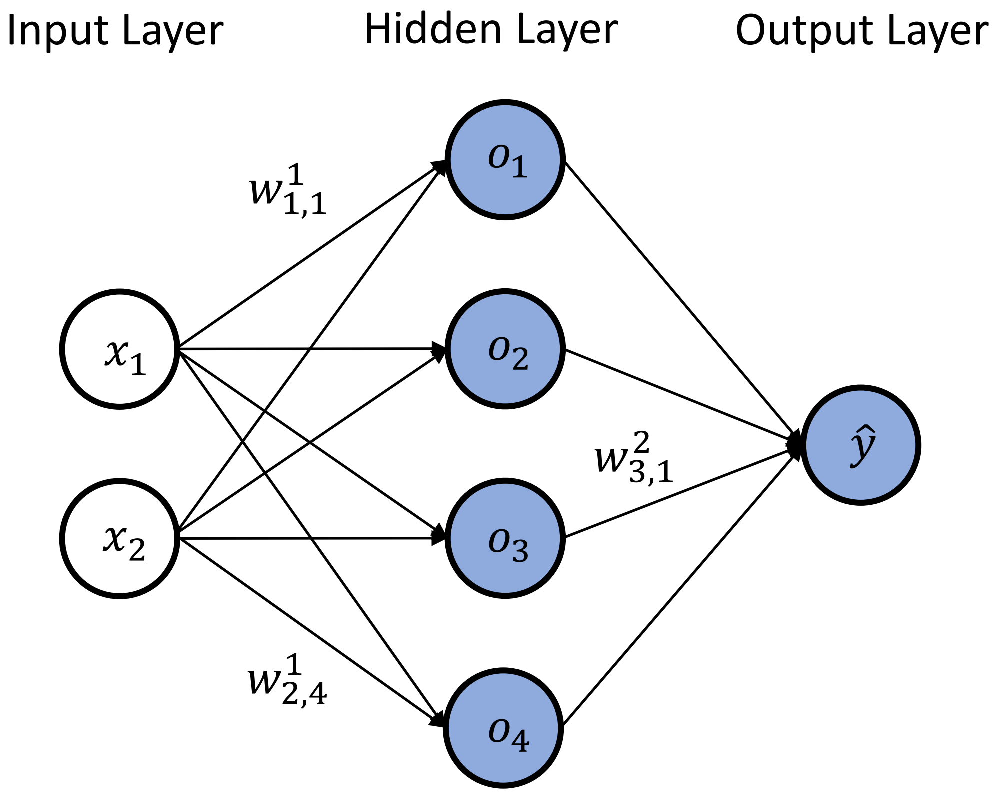
Income (Customer Feature 1)	Age (Customer Feature 2)	Revenue from Customer	Predicted Revenue Given Set of Weights W
22003	45	14.03875	$\hat{y}(22003,45;W)$
57230	54	23.31168	$\hat{y}(57230,54;W)$
75137	28	24.05046	$\hat{y}(75137,28;W)$
31208	54	18.5386	$\hat{y}(31208,54;W)$
54078	23	18.50195	$\hat{y}(54078,23;W)$
44413	44	20.63106	$\hat{y}(44413,44;W)$
55237	46	22.32953	$\hat{y}(55237,46;W)$



Training Dense Feed-Forward Neural Networks

Epoch #1, Step #1

Income (x_1)	Age (x_2)
22003	45
57230	54



Predicted Revenue Given Set of Weights W
$\hat{y}(22003{,}45 \mid W)$
$\hat{y}(57230{,}54 \mid W)$

Loss function is defined by comparing these

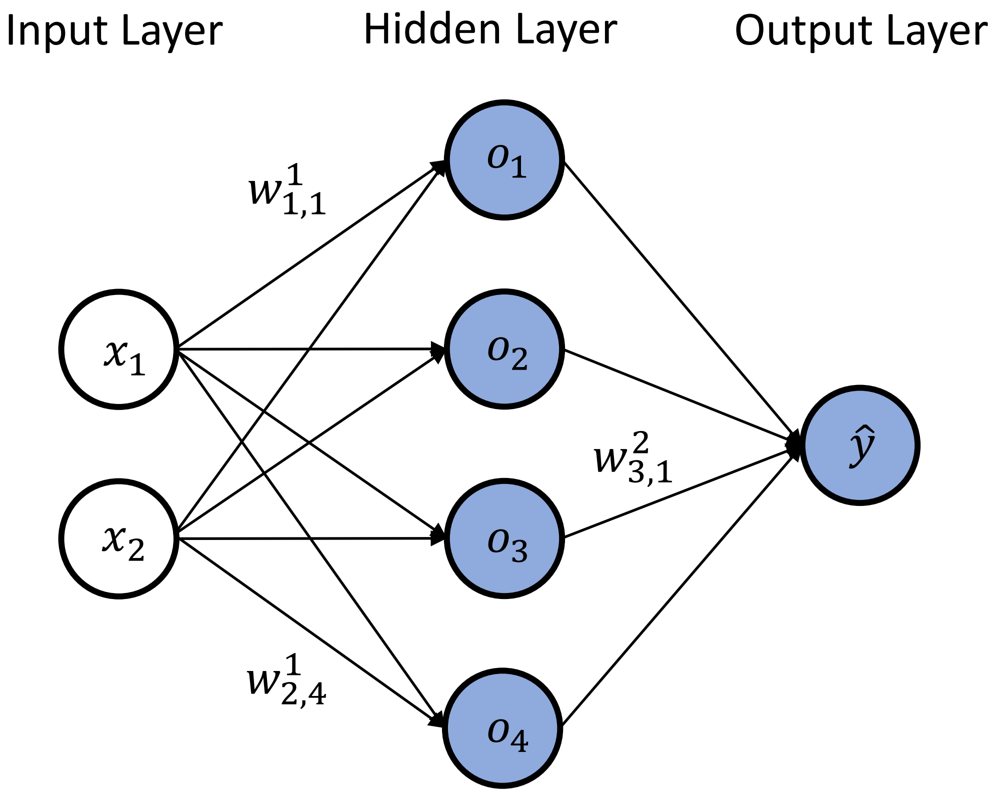
Revenue from Customer
14.03875
23.31168

Update weights based on loss function and optimizer

Training Dense Feed-Forward Neural Networks

Epoch #1, Step #2

Income (x_1)	Age (x_2)
75137	28
31208	54



Predicted Revenue Given Set of Weights W
$\hat{y}(75137,28 \mid W)$
$\hat{y}(31208,54 \mid W)$

Loss function is defined by comparing these

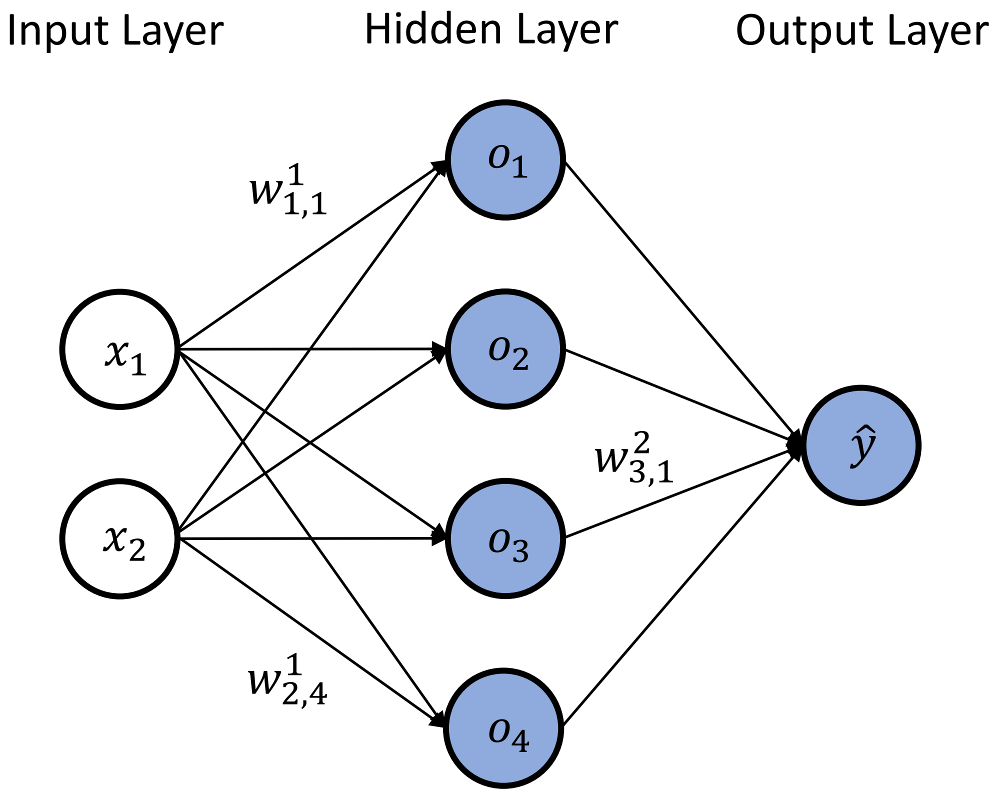
Revenue from Customer
24.05046
18.5386

Update weights based on loss function and optimizer

Training Dense Feed-Forward Neural Networks

Epoch #1, Step #3

Income (x_1)	Age (x_2)
54078	23
44413	44



Predicted Revenue Given Set of Weights W
$\hat{y}(54078,23 \mid W)$
$\hat{y}(44413,44 \mid W)$

Loss function is defined by comparing these

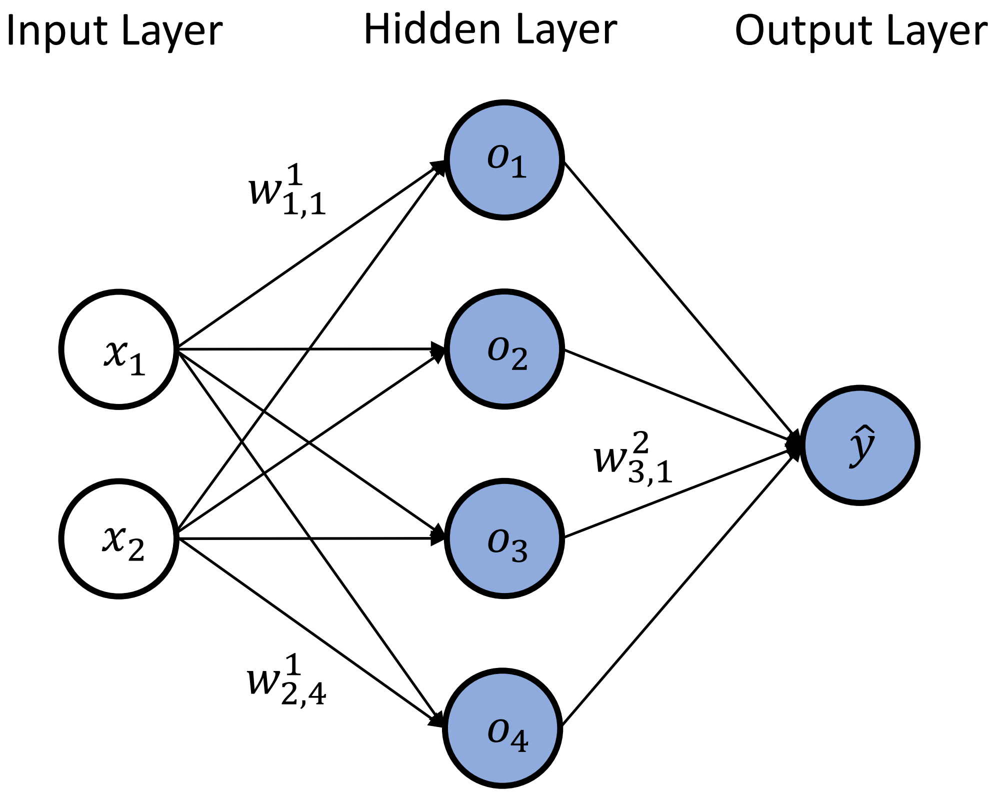
Revenue from Customer
18.50195
20.63106

Update weights based on loss function and optimizer

Training Dense Feed-Forward Neural Networks

Epoch #1, Step #4

Income (x_1)	Age (x_2)
55237	46



Predicted Revenue Given Set of Weights W
$\hat{y}(55237{,}46 \mid W)$

Loss function is defined by comparing these

Revenue from Customer
22.32953

Update weights based on loss function and optimizer

Training Dense Feed-Forward Neural Networks

- Finished iterating through the training set
    - Epoch #1 is now complete
- Shuffle the training set and start Epoch #2
- Continue in same manner until all epochs are completed

Income (Customer Feature 1)	Age (Customer Feature 2)	Revenue from Customer
22003	45	14.03875
57230	54	23.31168
75137	28	24.05046
31208	54	18.5386
54078	23	18.50195
44413	44	20.63106
55237	46	22.32953