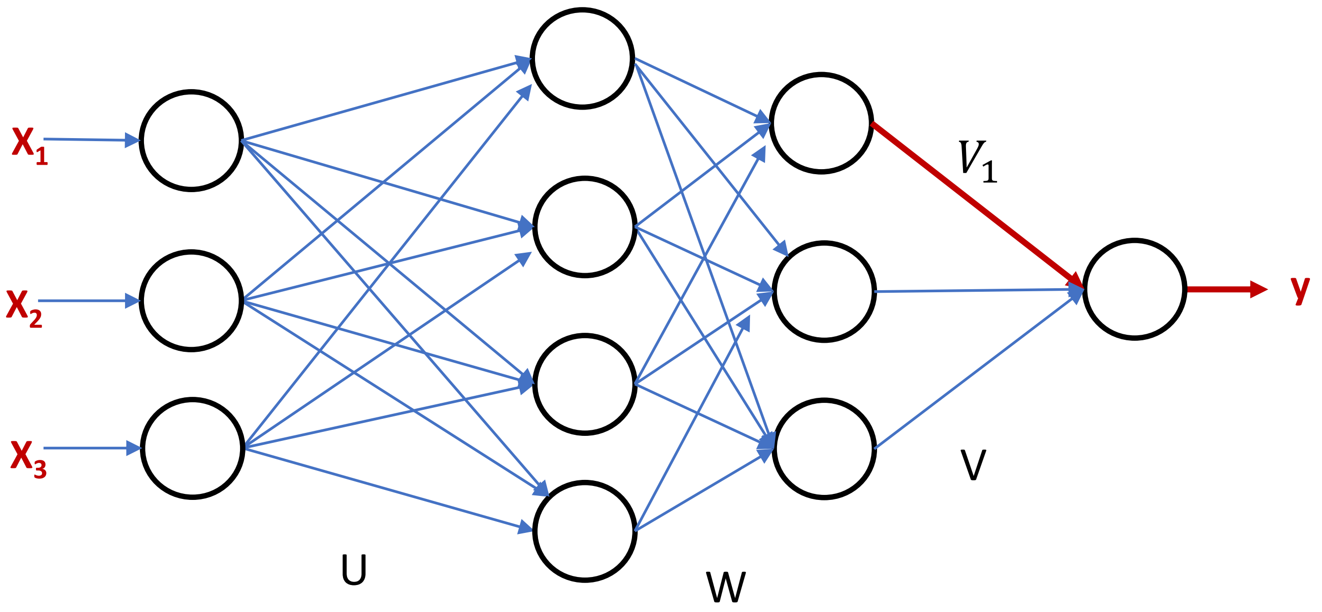
Lesson 11

Learning with different Loss Functions and Their Derivatives

Two Commonly used Loss Functions are

- Mean Square Error – Standard Loss Function for Regression
- Cross Entropy Loss - Standard Loss Function for Classification

Mean Square Error (MSE)



For the Single Sample

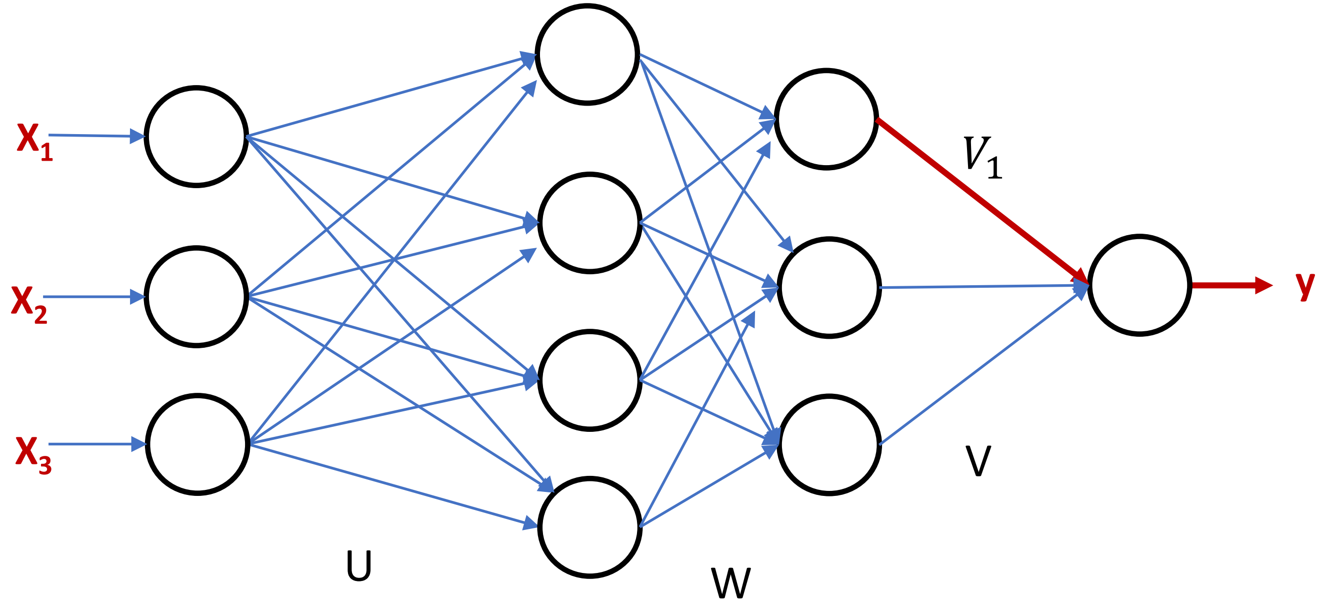
MSE $E = (y - \acute{y})^2$

Ground truth is $\acute{y}$

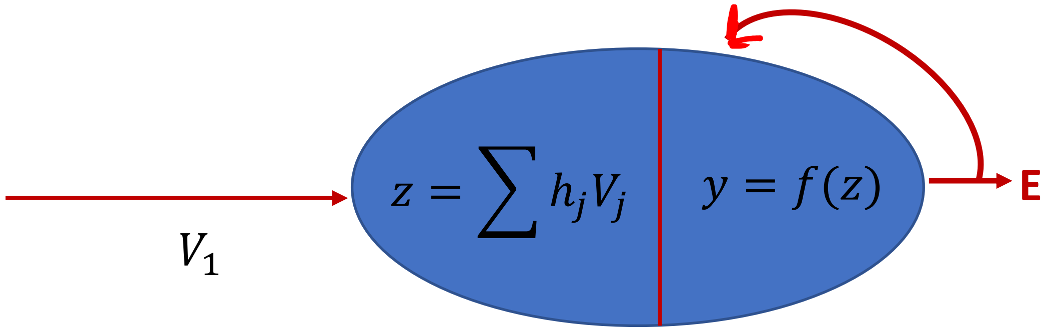
Mean Square Error (MSE)

For the Single Sample

MSE $E = (y - \acute{y})^2$

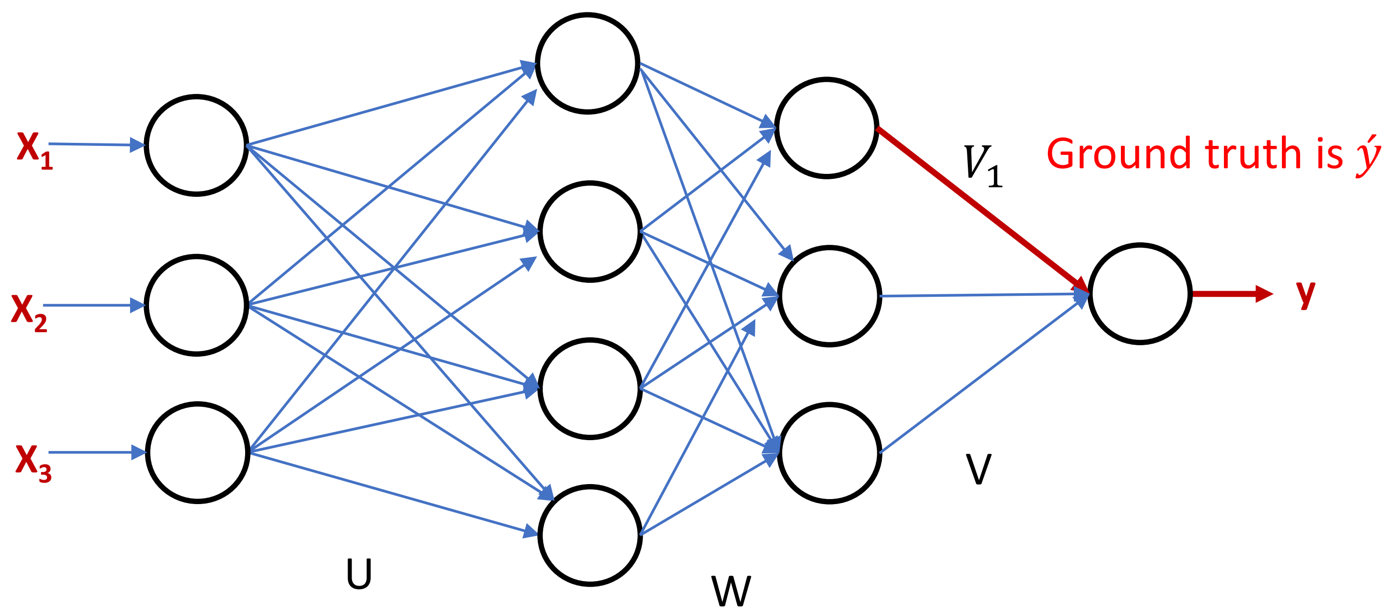


Ground truth is $\acute{y}$



$$\frac{\delta E}{\delta V_1} = \frac{\delta z}{\delta V_1} \times \frac{\delta y}{\delta z} \times \frac{\delta E}{\delta y}$$

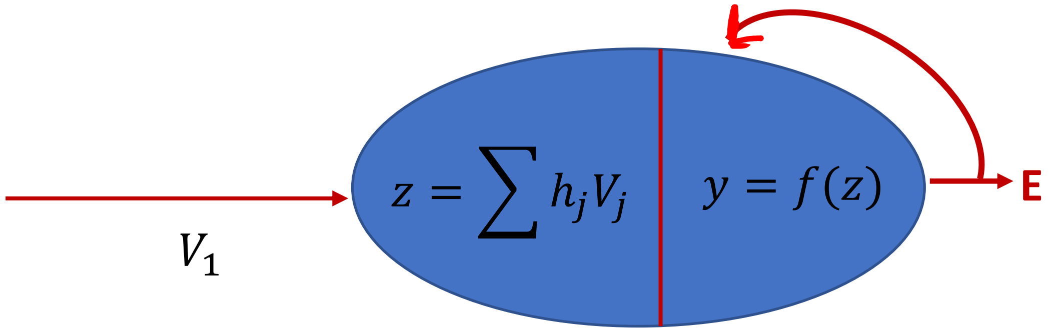
Mean Square Errors



For the Single Sample

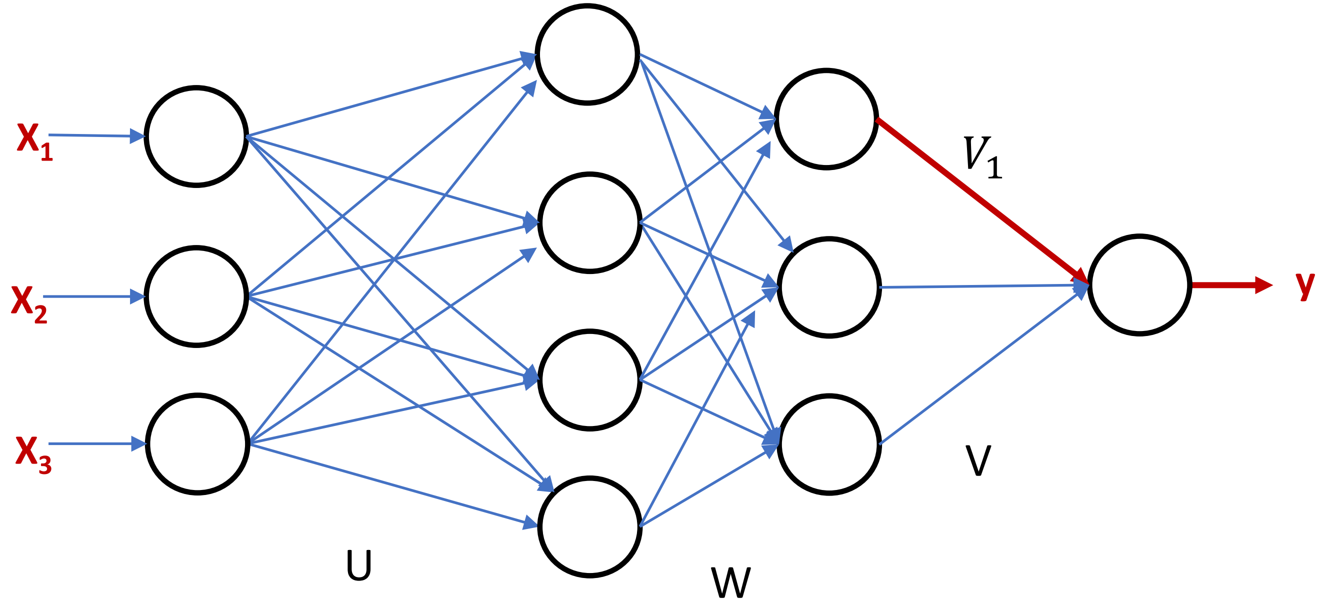
$$E = (y - \acute{y})^2$$

$$\frac{\delta E}{\delta y} = \frac{\delta (y - \acute{y})^2}{\delta y} = 2(y - \acute{y})$$



$$\frac{\delta E}{\delta V_1} = \frac{\delta z}{\delta V_1} \times \frac{\delta y}{\delta z} \times \frac{\delta E}{\delta y}$$

Mean Square Errors

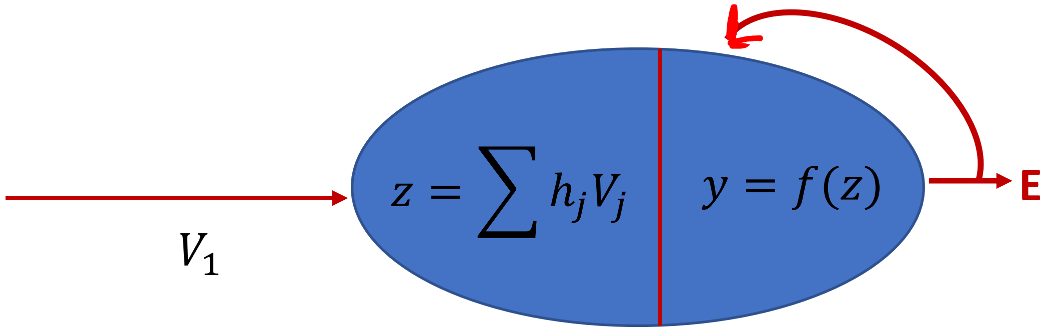


If the ground truth is $\acute{y}$

For *n* Samples

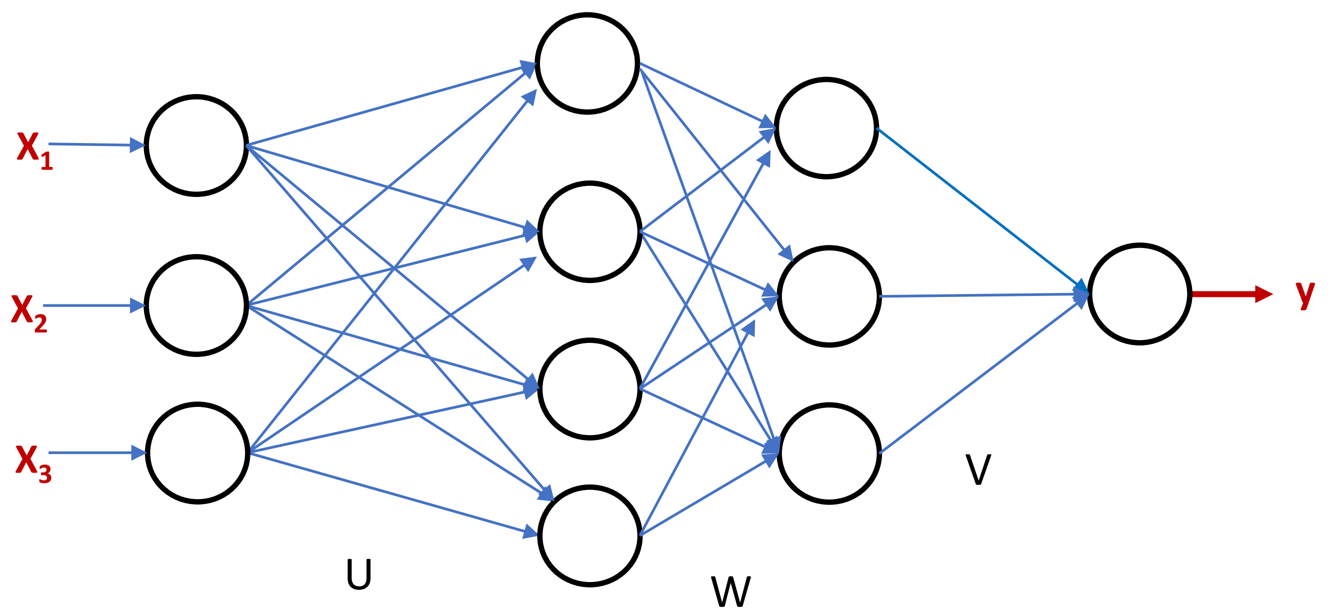
$$E = \frac{1}{n}\sum_{i=1}^{n}(y^i - \acute{y}^i)^2$$

$$\frac{\delta E}{\delta y} = \frac{\delta \frac{1}{n}\sum_{i=1}^{n}(y^i - \acute{y}^i)^2}{\delta y} = \frac{2}{n}\sum_{i=1}^{n}(y^i - \acute{y}^i)$$



$$\frac{\delta E}{\delta V_1} = \frac{\delta z}{\delta V_1} \times \frac{\delta y}{\delta z} \times \frac{\delta E}{\delta y}$$

Mean Square Errors



If the ground truth is $\acute{y}$

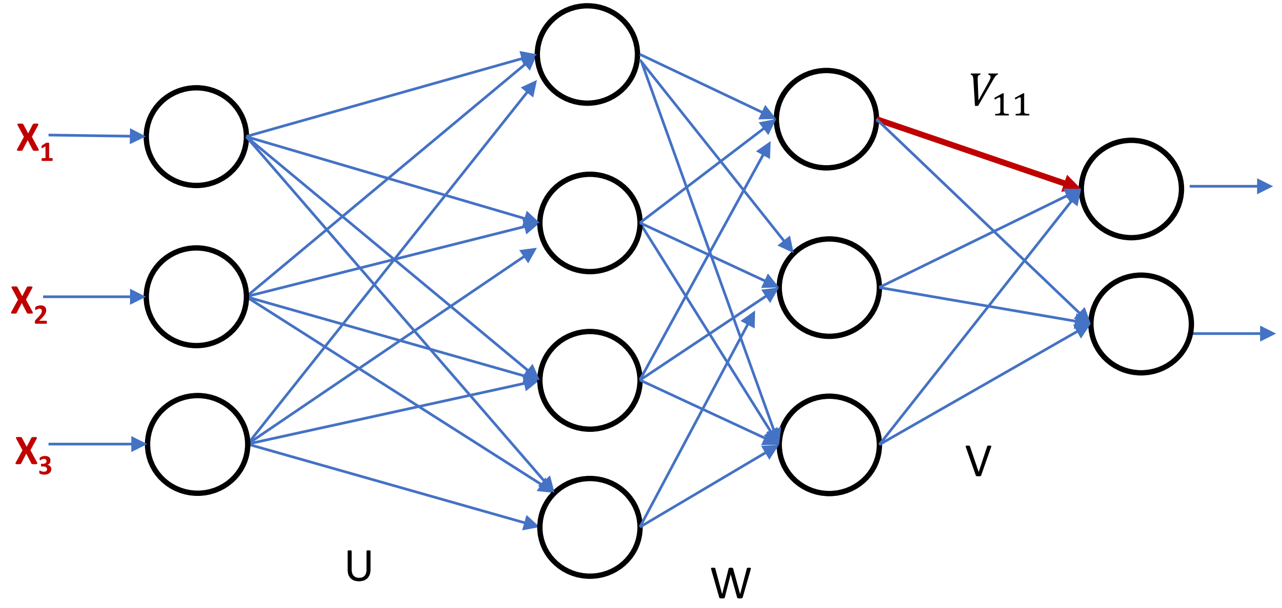
For *n* Samples

$$E = \frac{1}{n}\sum_{i=1}^{n}(y^i - \acute{y}^i)^2$$

$$\frac{\delta E}{\delta y} = \frac{\delta \frac{1}{n}\sum_{i=1}^{n}(y^i - \acute{y}^i)^2}{\delta y} = \frac{2}{n}\sum_{i=1}^{n}(y^i - \acute{y}^i)$$

Backpropagation will be done after a batch of *n* Samples

Mean Square Errors

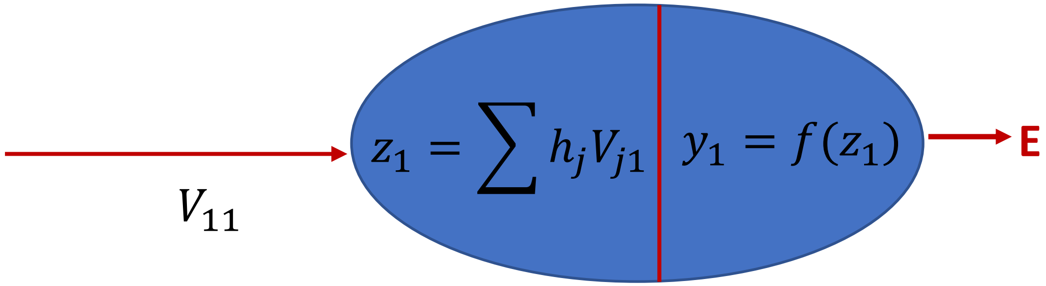


If the ground truth is $\acute{y}$

For the n Sample

$$\boldsymbol{E} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{y}^i - \acute{\boldsymbol{y}}^i)^2$$

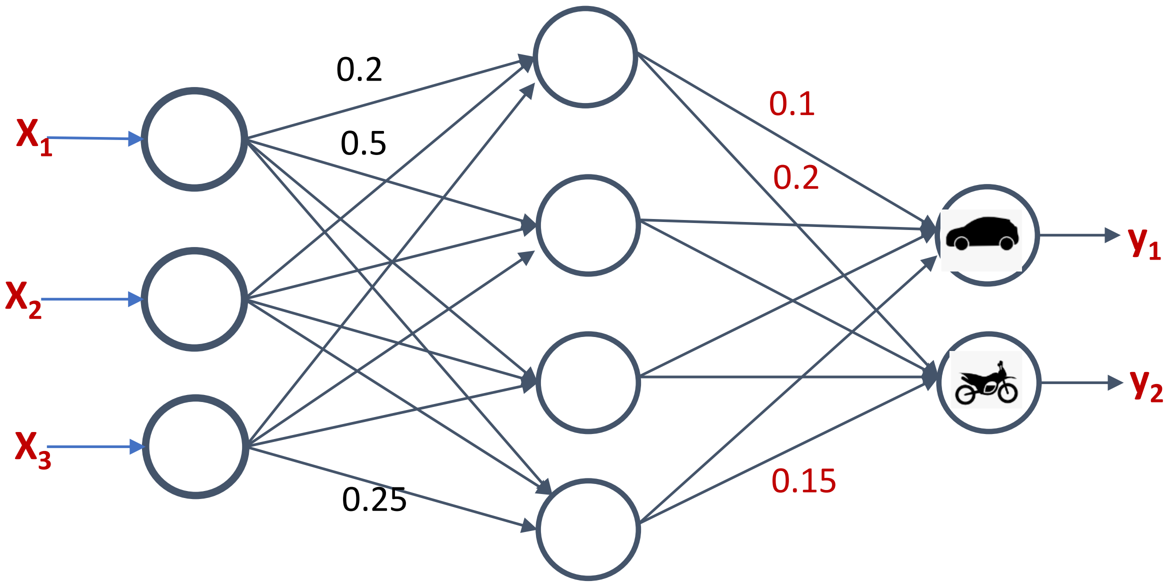
$$\frac{\delta \boldsymbol{E}}{\delta y_1} = \frac{\delta \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{y}^i - \acute{\boldsymbol{y}}^i)^2}{\delta y_1} = \frac{2}{n}\sum_{i=1}^{n}({y_1}^i - {\acute{y}_1}^i)$$



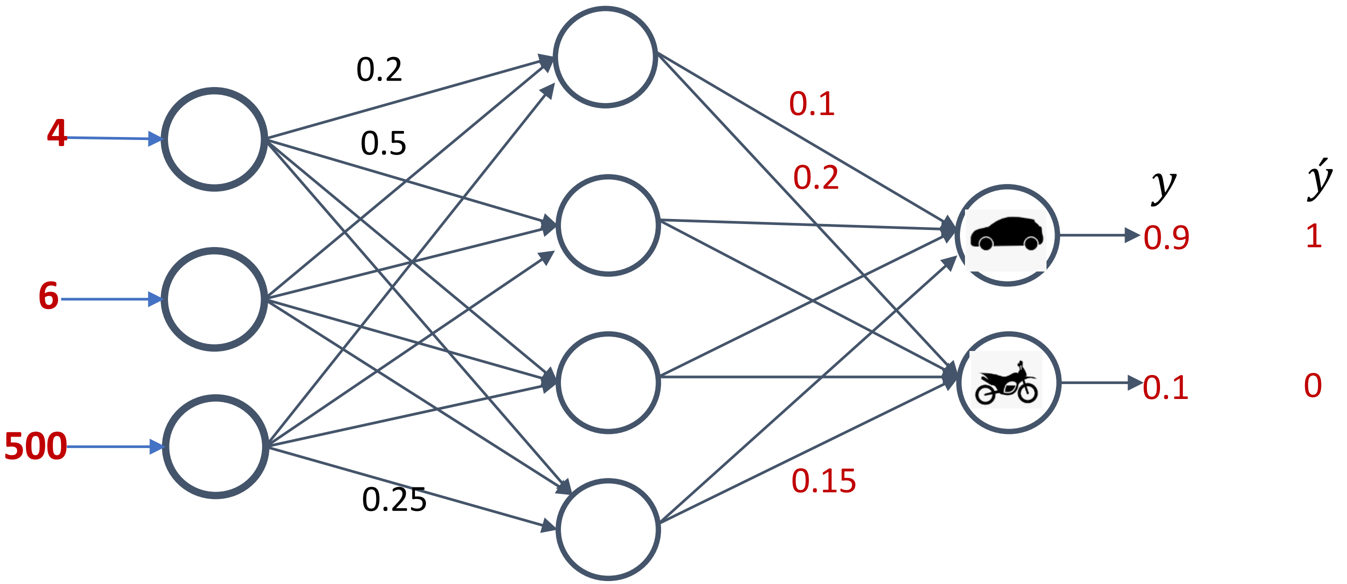
$$\frac{\delta E}{\delta V_{11}} = \frac{\delta z_1}{\delta V_{11}} \times \frac{\delta y_1}{\delta z_1} \times \frac{\delta \boldsymbol{E}}{\delta y_1}$$

Let us illustrate with a toy example

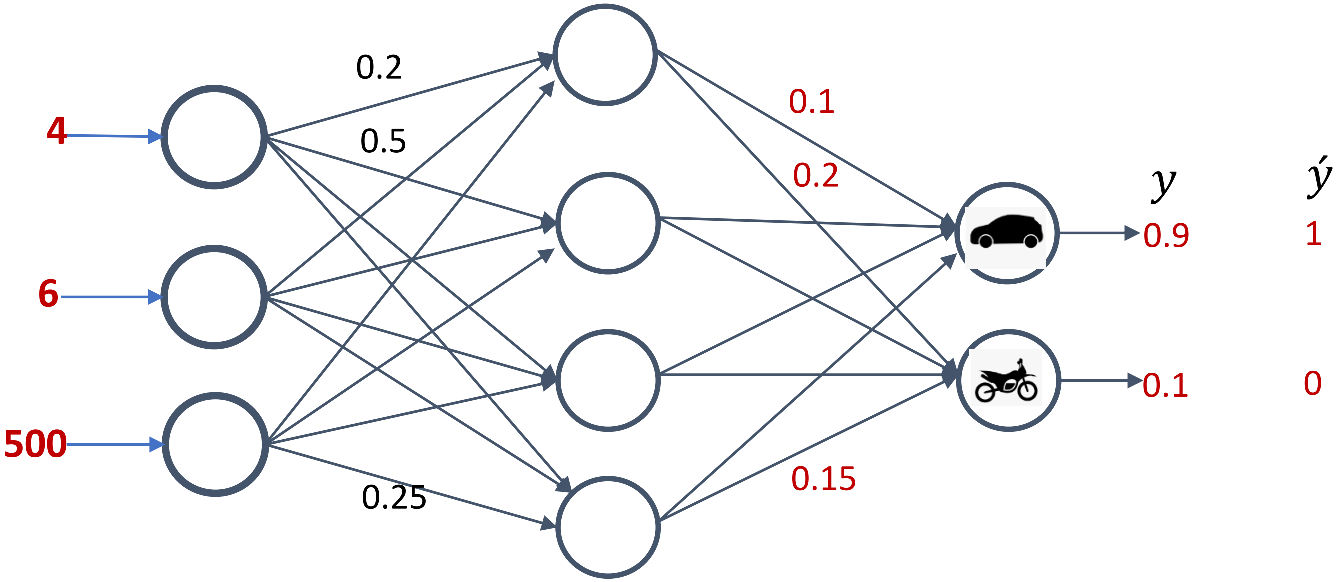
#Wheel	Height	Weight	
4	6	500	
4	5.5	600	
4	5	550	
2	3	200	
2	3.5	150	
2	4	250	



#Wheel	Height	Weight
4	6	500
4	5.5	600
4	5	550
2	3	200
2	3.5	150
2	4	250



#Wheel	Height	Weight	
4	6	500	
4	5.5	600	
4	5	550	
2	3	200	
2	3.5	150	
2	4	250	



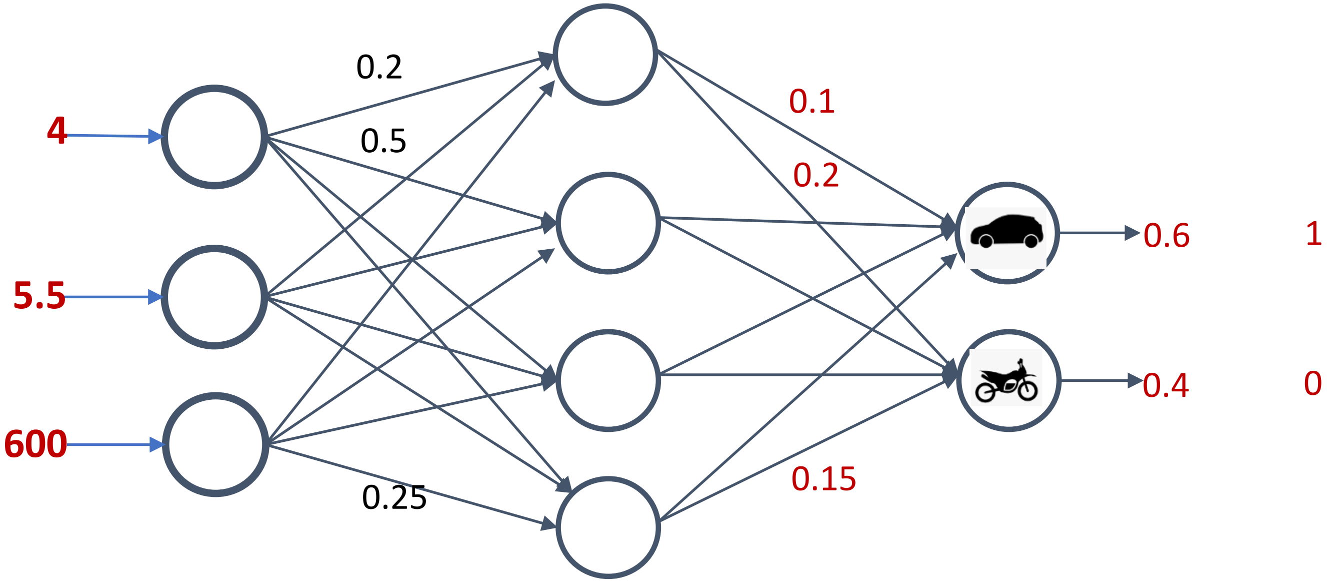
$$\text{Error } E = (y - \acute{y})^2$$

$$\text{Error } E_{y_1} = (0.9 - 1)^2 = 0.01$$

$$\text{Error } E_{y_2} = (0.1 - 0)^2 = 0.01$$

If these errors are not acceptable, then Backpropagate.

#Wheel	Height	Weight	
4	6	500	
4	5.5	600	
4	5	550	
2	3	200	
2	3.5	150	
2	4	250	



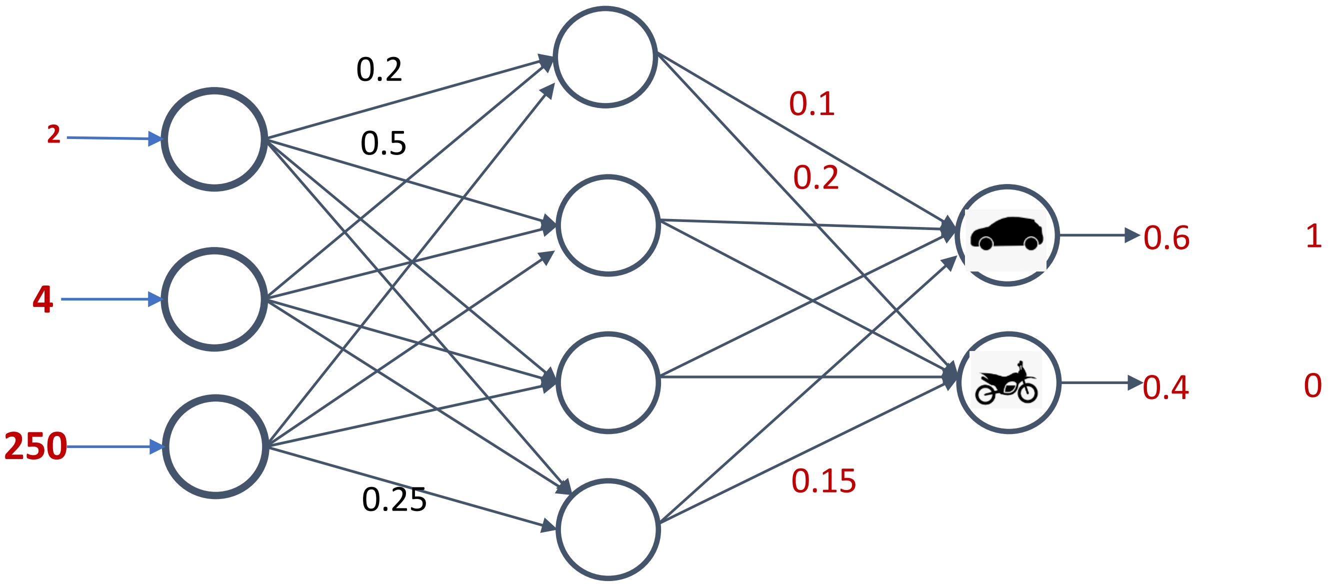
Error $E = (y - \acute{y})^2$

Error $E_{y_1} = (0.6 - 1)^2 = 0.16$

Error $E_{y_2} = (0.4 - 0)^2 = 0.36$

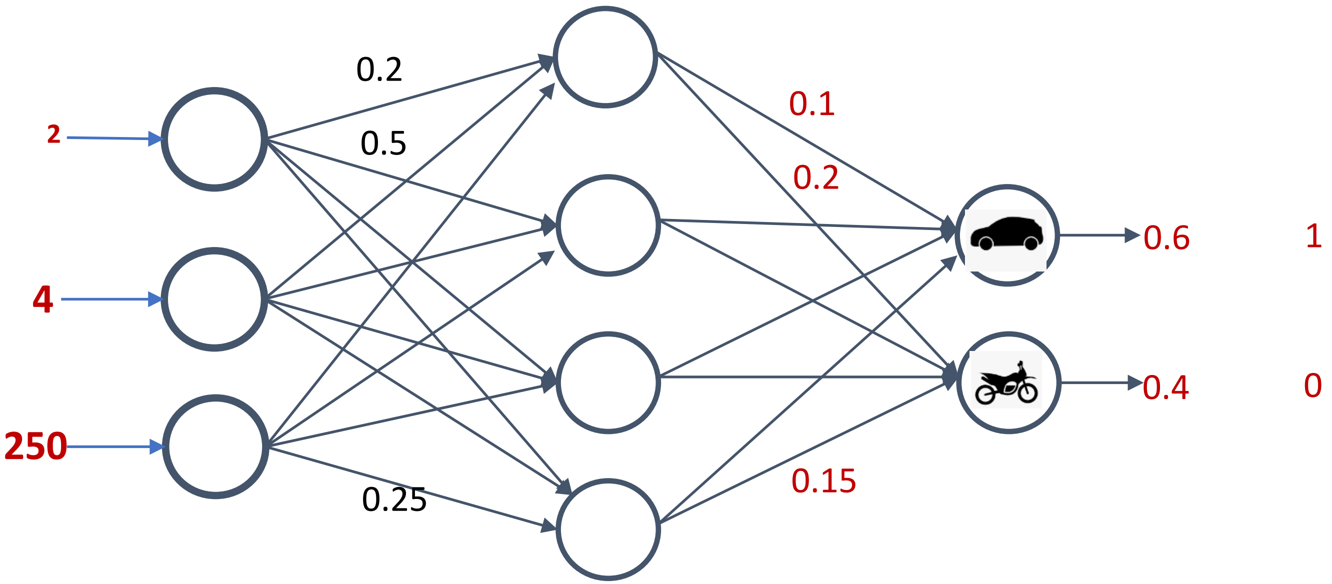
If these errors are not acceptable, then Backpropagate.

#Wheel	Height	Weight	
4	6	500	
4	5.5	600	
4	5	550	
2	3	200	
2	3.5	150	
2	4	250	



One complete cycle of training is called Epoch

#Wheel	Height	Weight	
4	6	500	
4	5.5	600	
4	5	550	
2	3	200	
2	3.5	150	
2	4	250	



Backpropagation after every sample is expensive.

Do it in batches.

Summary

- Mean Square Loss Function and how to estimate its gradient