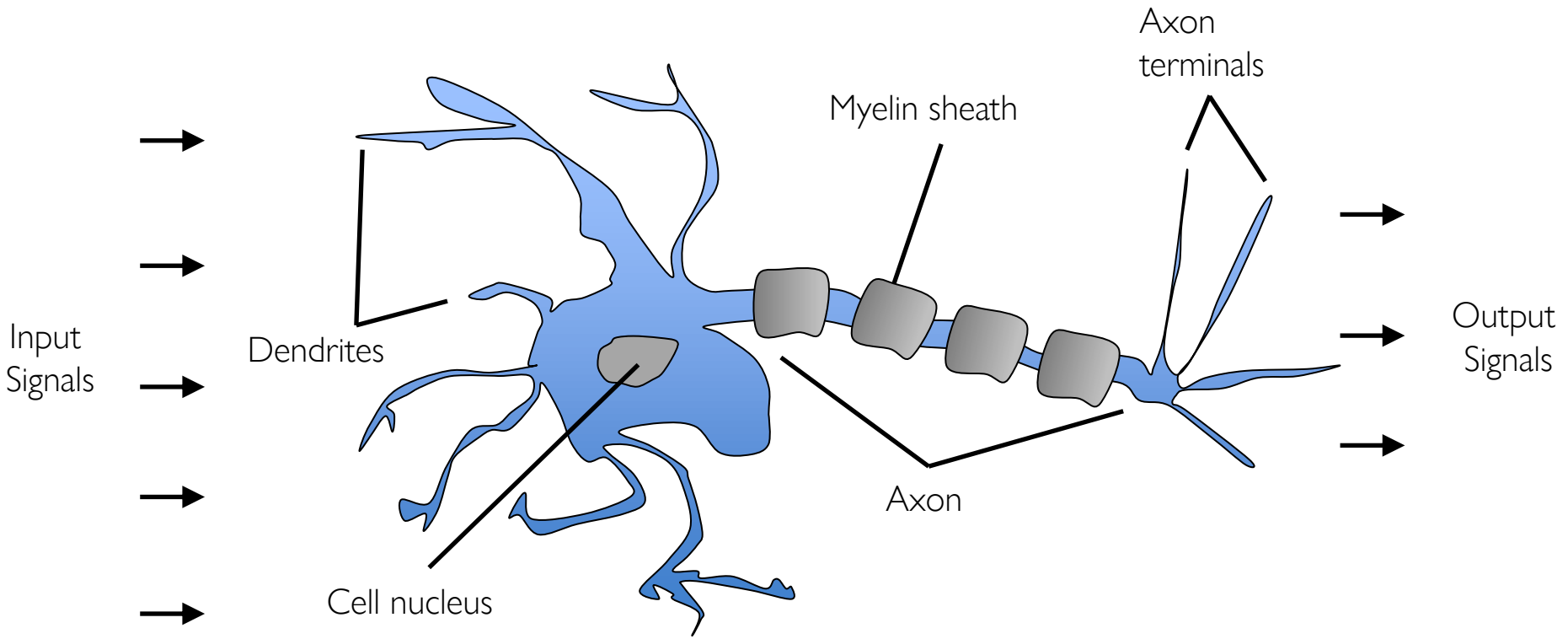
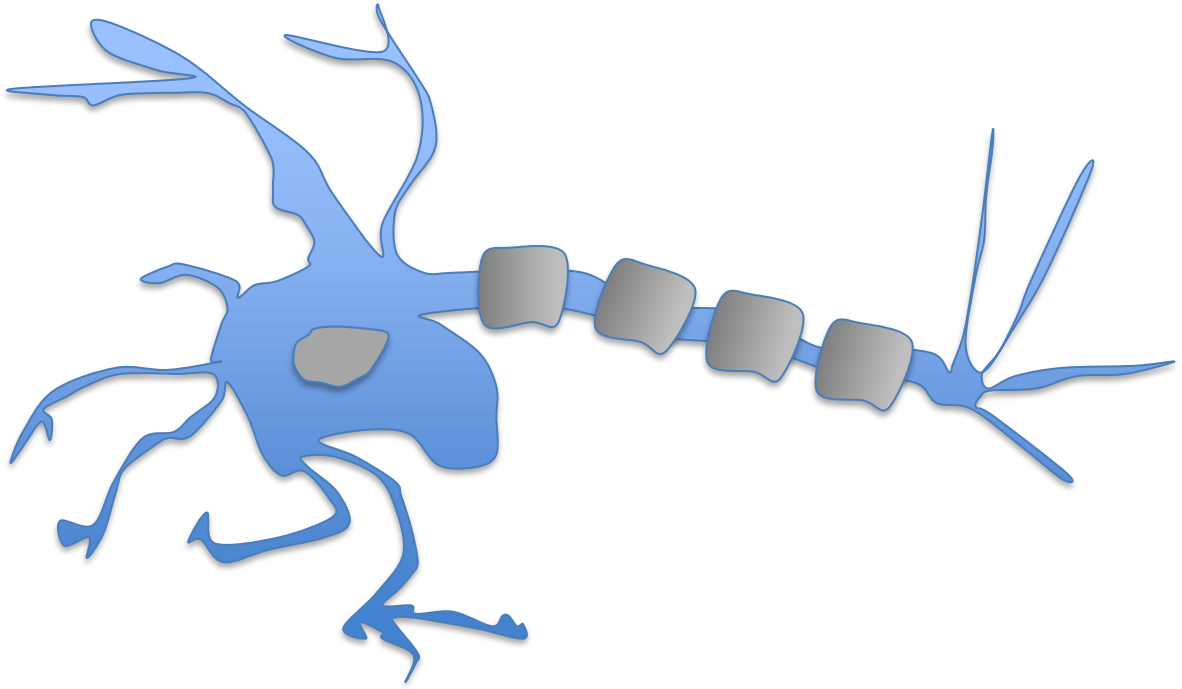
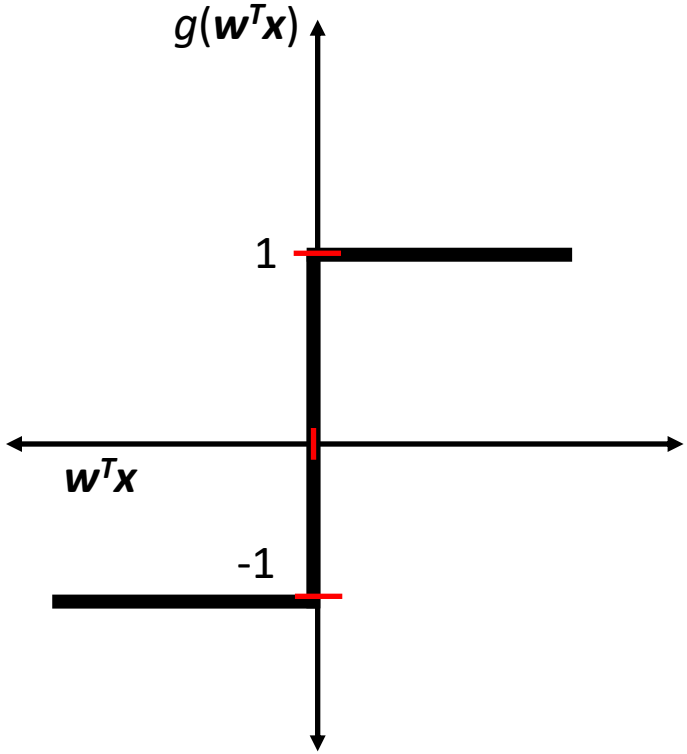
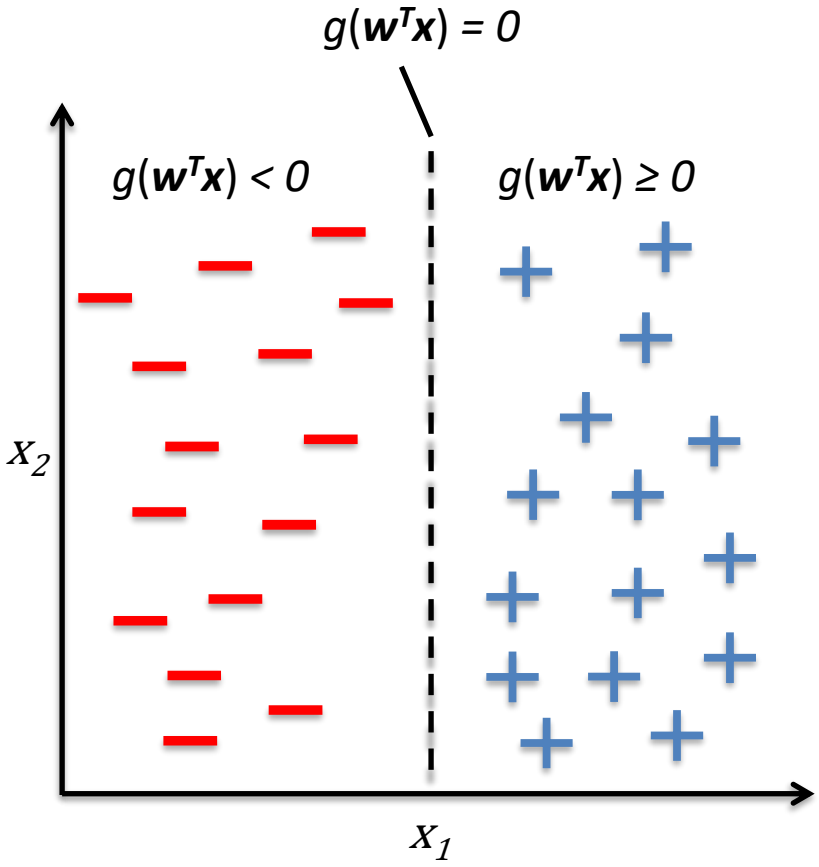
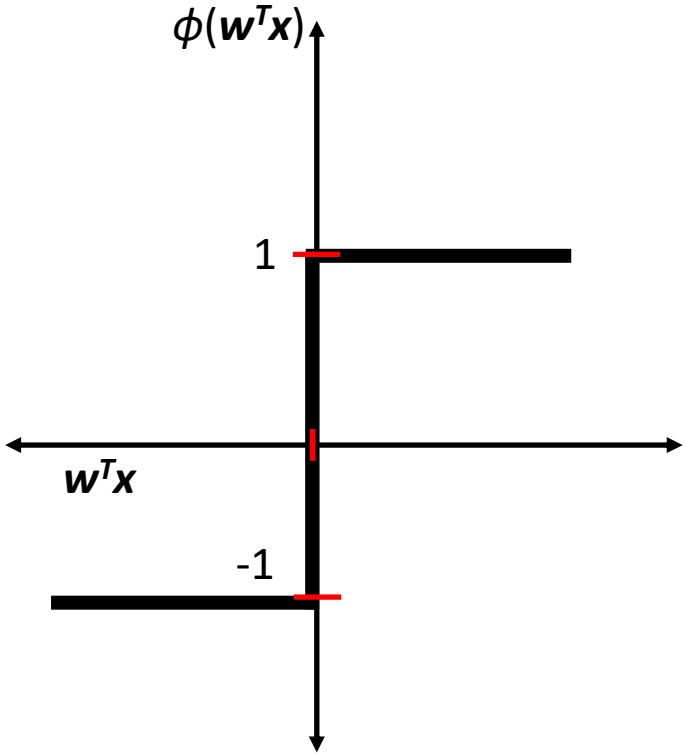


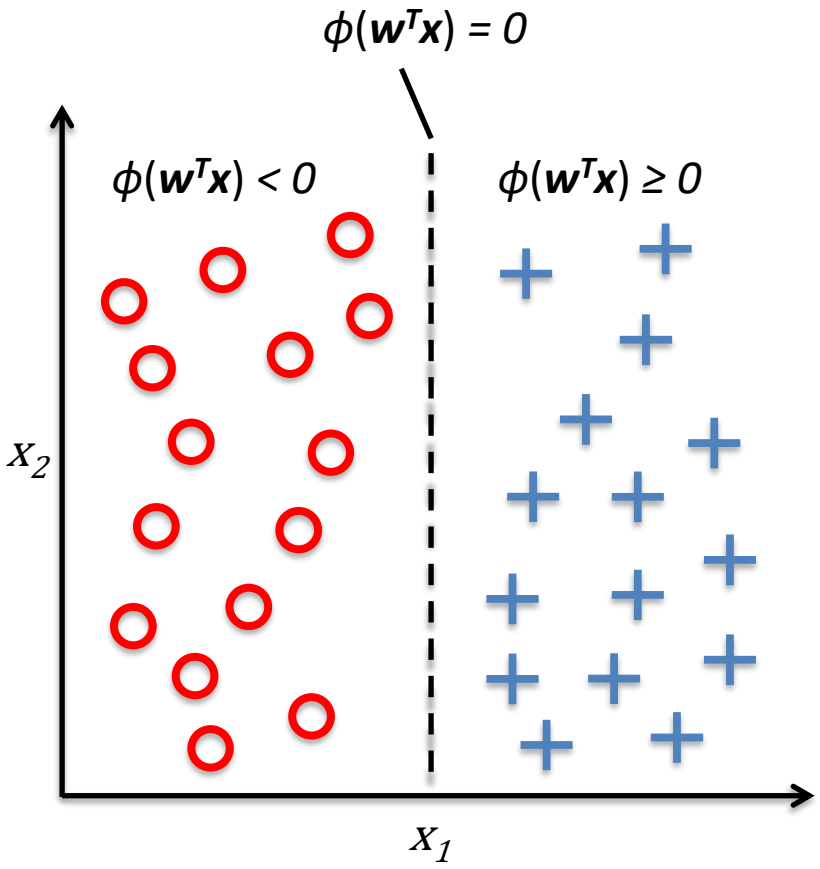
Schematic of a biological neuron.

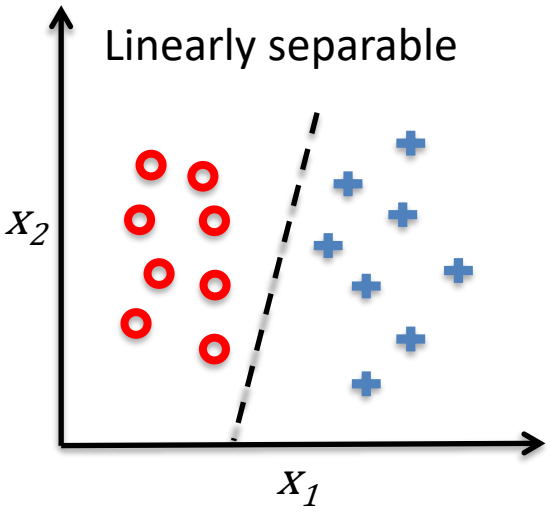


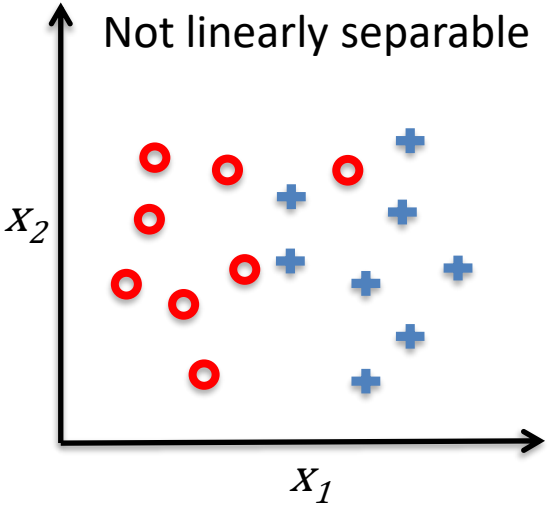

$g(\mathbf{w}^T\mathbf{x})$
1
$\mathbf{w}^T\mathbf{x}$
-1

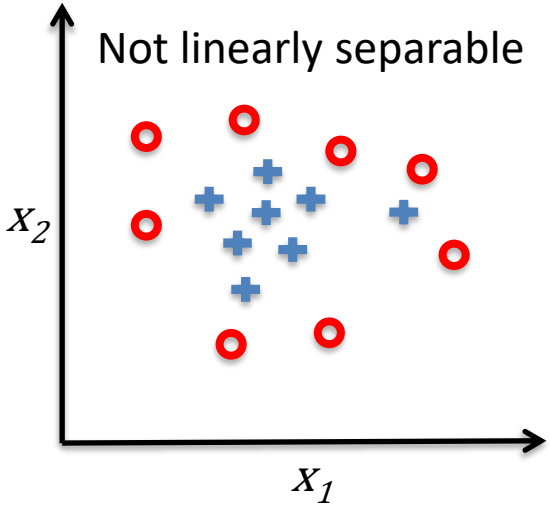

$g(\mathbf{w}^T\mathbf{x}) = 0$
$g(\mathbf{w}^T\mathbf{x}) < 0$
$g(\mathbf{w}^T\mathbf{x}) \geq 0$
x_2
x_1

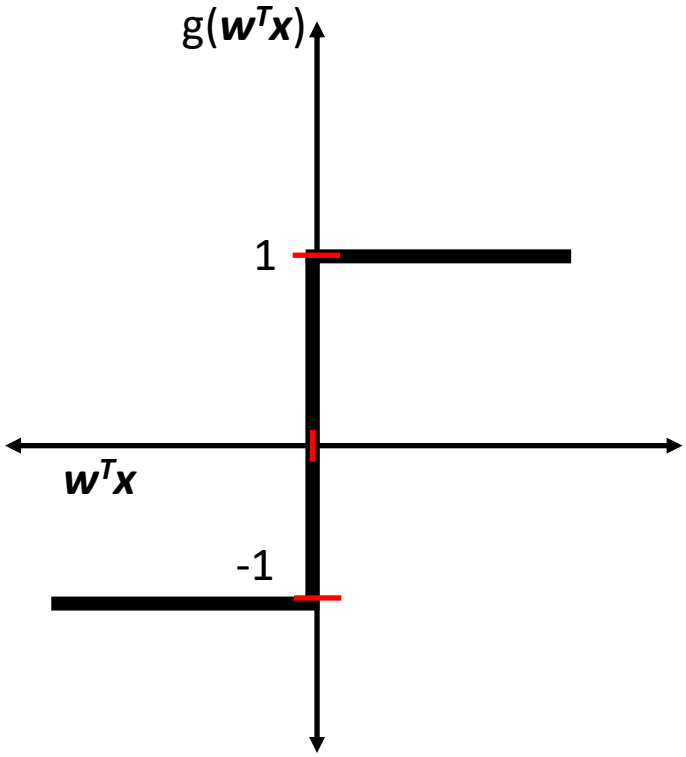

$\phi(\boldsymbol{w}^T\boldsymbol{x})$
1
$\boldsymbol{w}^T\boldsymbol{x}$
-1


$\phi(\boldsymbol{w}^T\boldsymbol{x}) = 0$
$\phi(\boldsymbol{w}^T\boldsymbol{x}) < 0$
$\phi(\boldsymbol{w}^T\boldsymbol{x}) \geq 0$
x_2
x_1

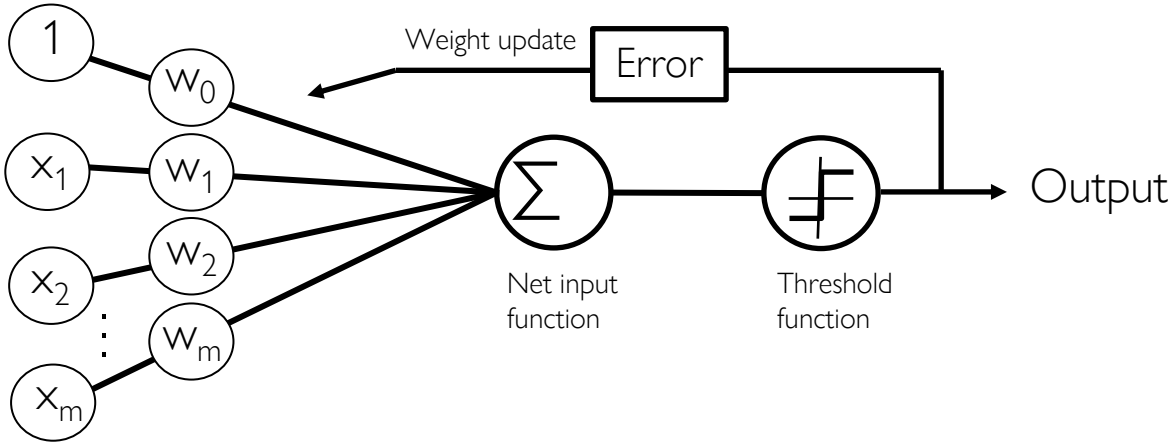

Linearly separable
X_2
X_1

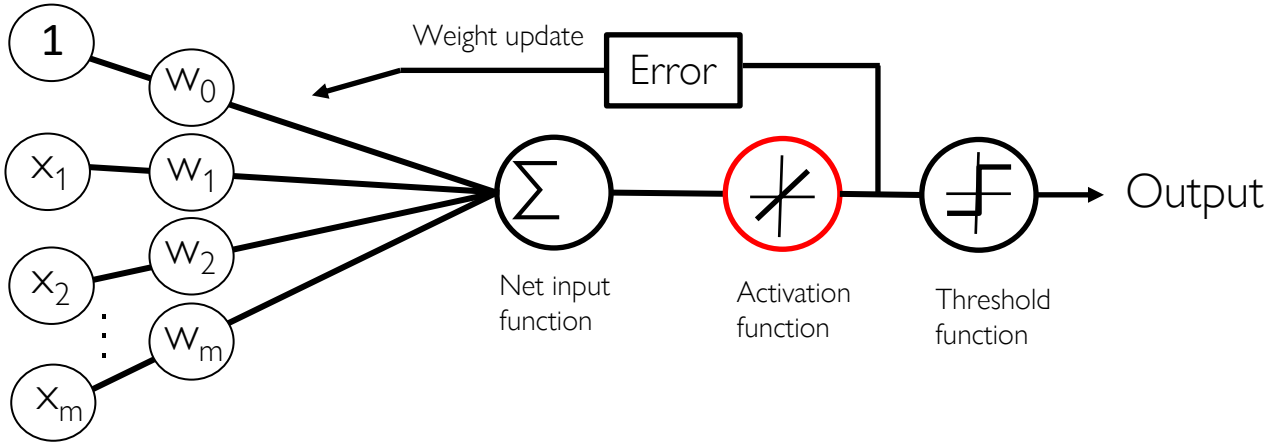

Not linearly separable
X_2
X_1


Not linearly separable
X_2
X_1

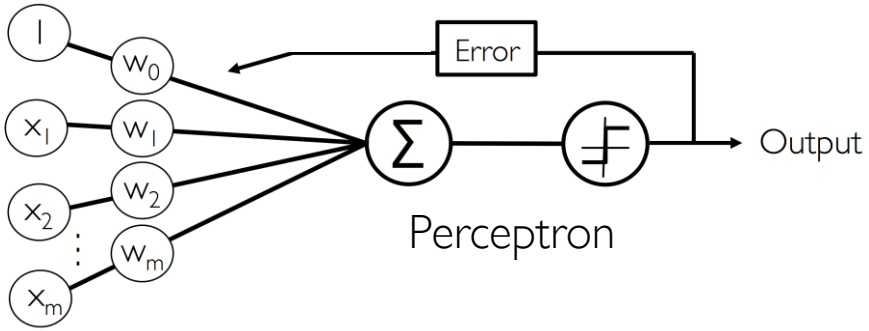


Unit step function.

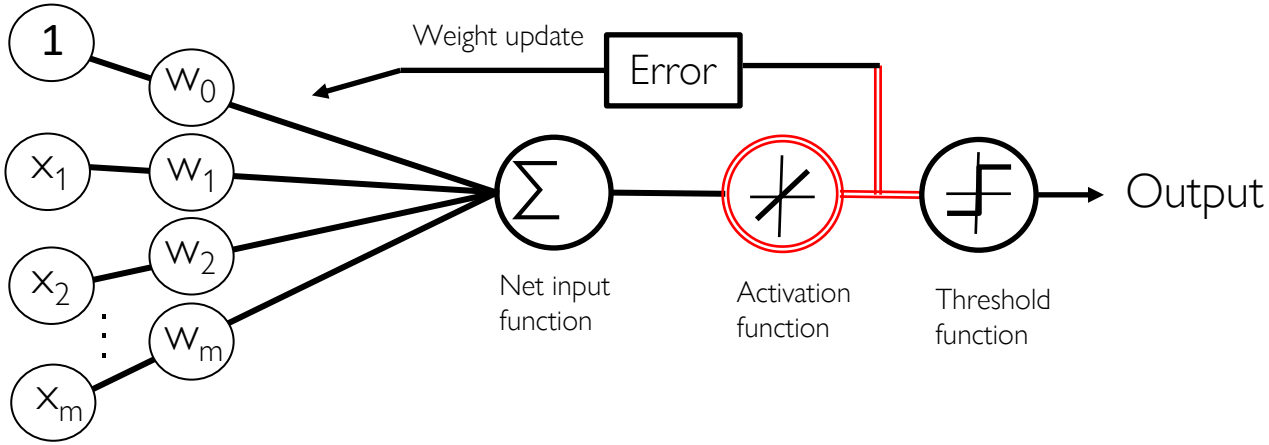




Adaptive Linear Neuron (Adaline)

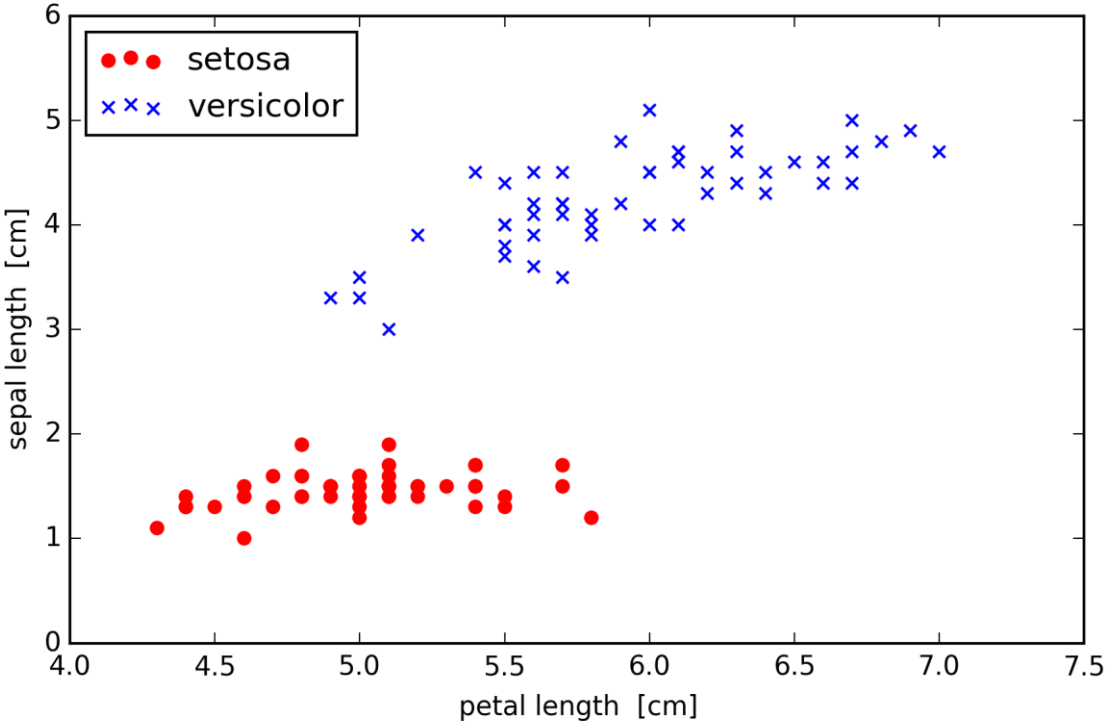


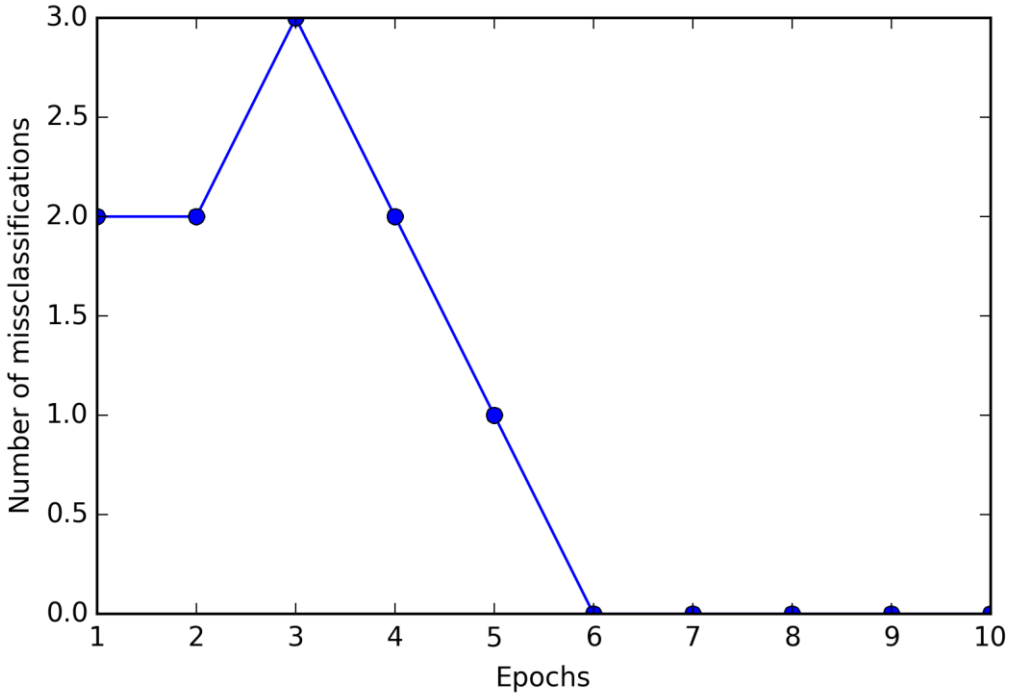
Perceptron

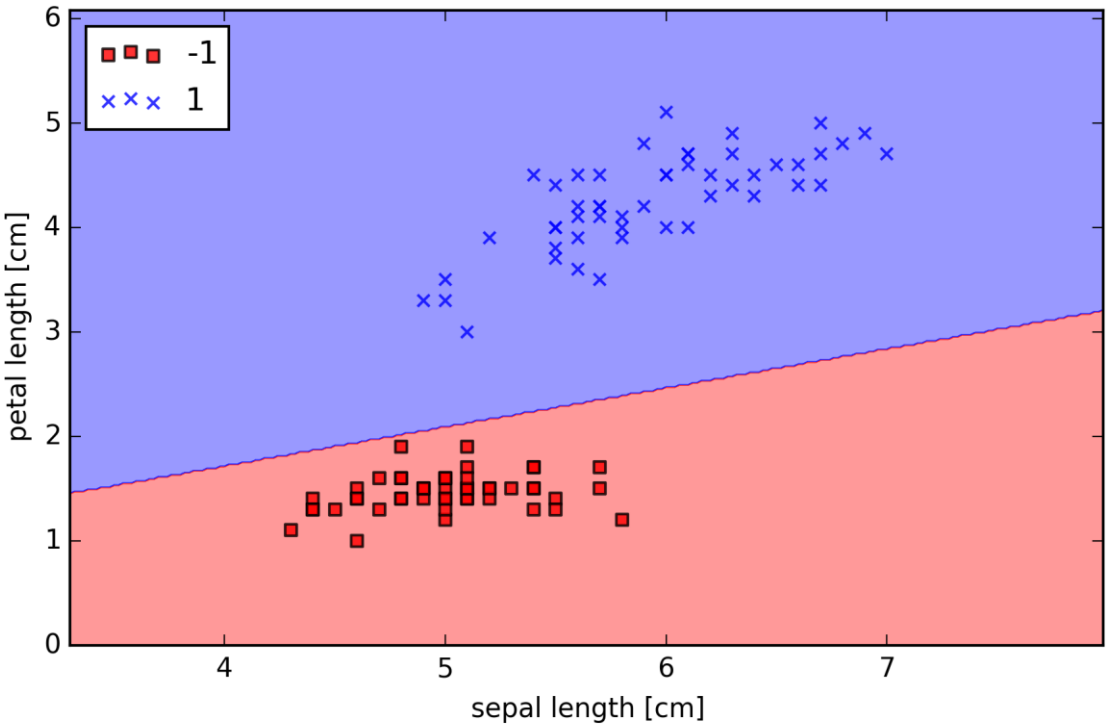


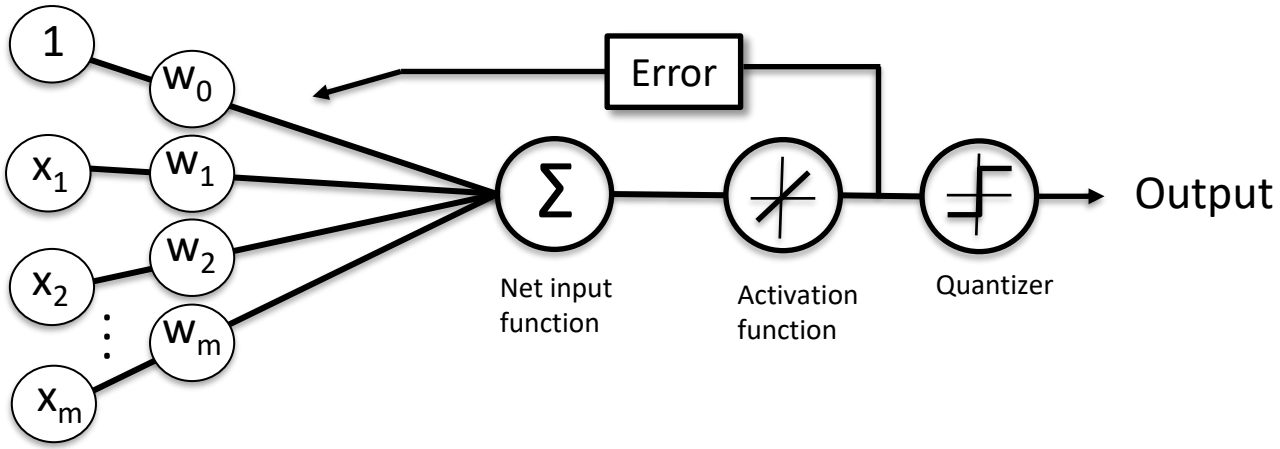
Adaptive Linear Neuron (Adaline)

	0	1	2	3	4
145	6.7	3.0	5.2	2.3	Iris-virginica
146	6.3	2.5	5.0	1.9	Iris-virginica
147	6.5	3.0	5.2	2.0	Iris-virginica
148	6.2	3.4	5.4	2.3	Iris-virginica
149	5.9	3.0	5.1	1.8	Iris-virginica

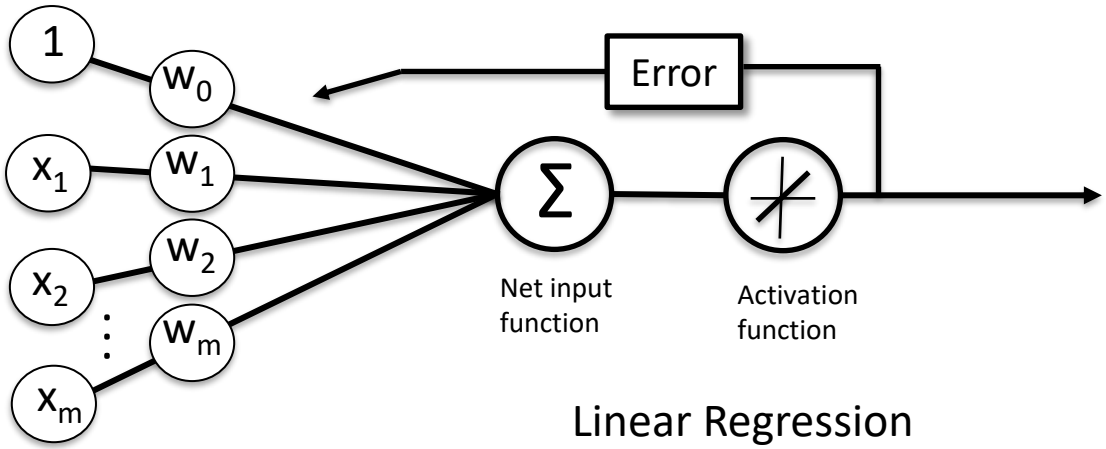

setosa
versicolor
sepal length [cm]
petal length [cm]
0
1
2
3
4
5
6
4.0
4.5
5.0
5.5
6.0
6.5
7.0
7.5


3.0
2.5
2.0
1.5
1.0
0.5
0.0
Number of missclassifications
1
2
3
4
5
6
7
8
9
10
Epochs

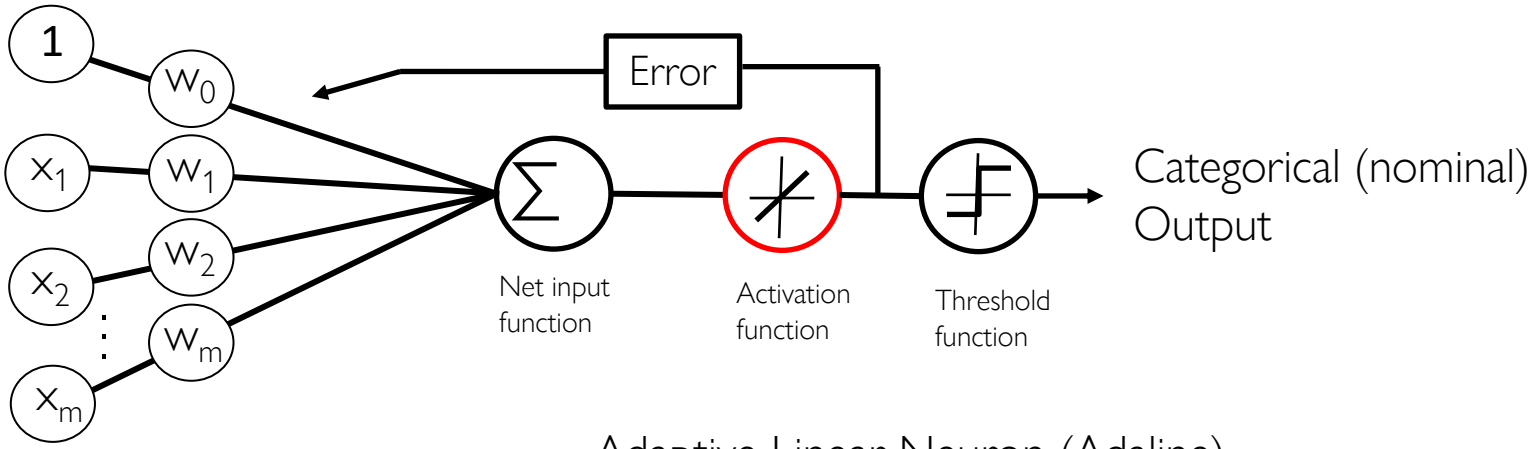

-1
1
petal length [cm]
sepal length [cm]
0
1
2
3
4
5
6
4
5
6
7


1
w_0
x_1
w_1
w_2
x_2
w_m
x_m
Error
Σ
Net input function
Activation function
Quantizer
Output

Adaline.



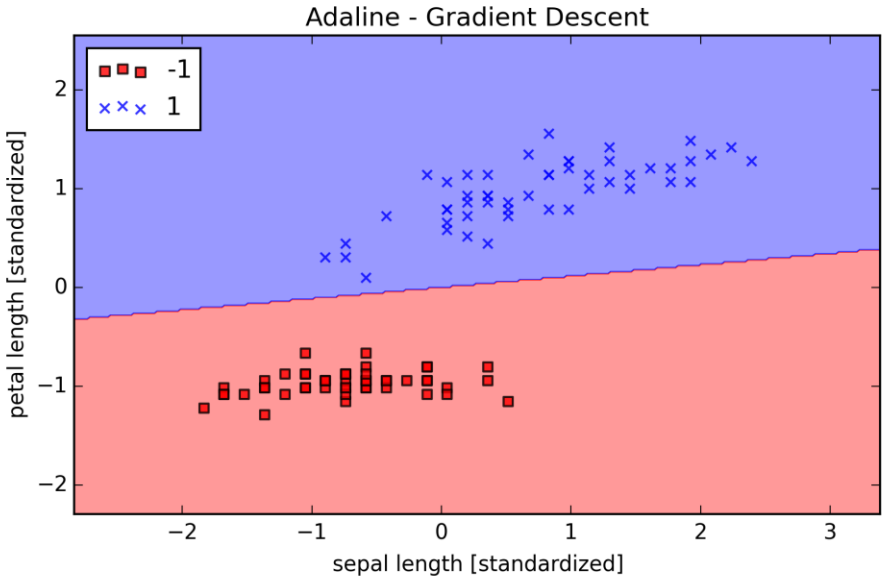
Linear Regression

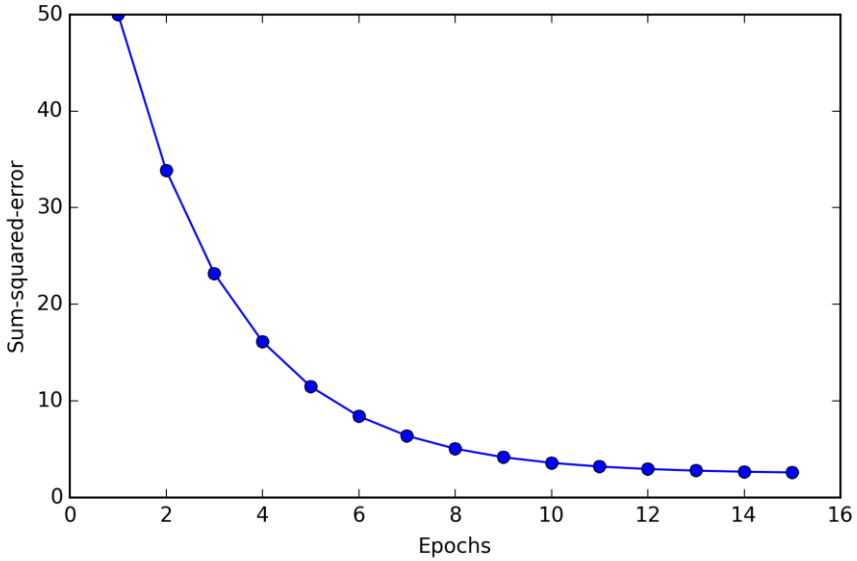


Adaptive Linear Neuron (Adaline)

iris_adaline_gd_cost_convergence_1.png

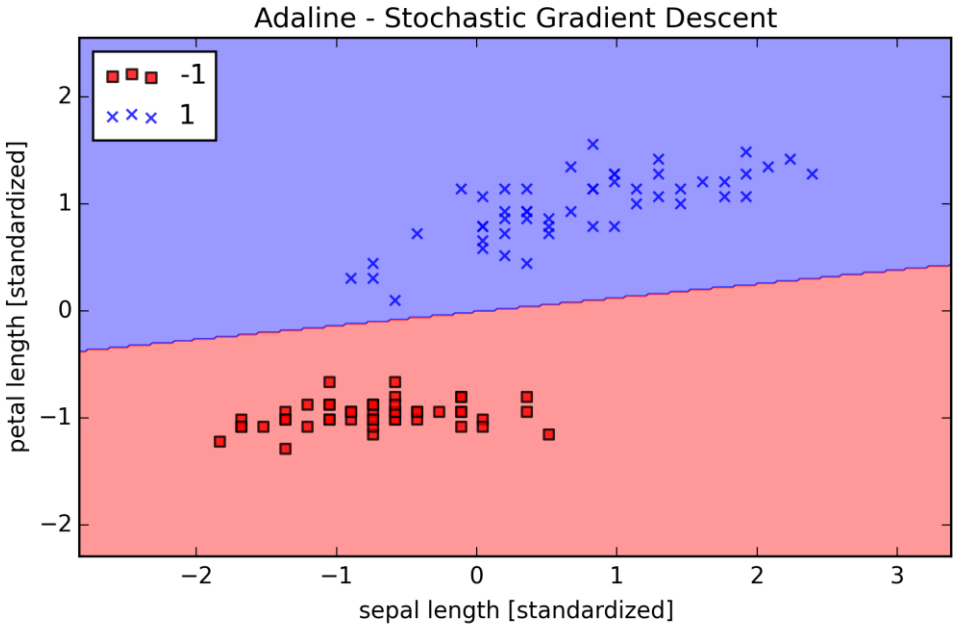
iris_adaline_gd_cost_convergence_2.png

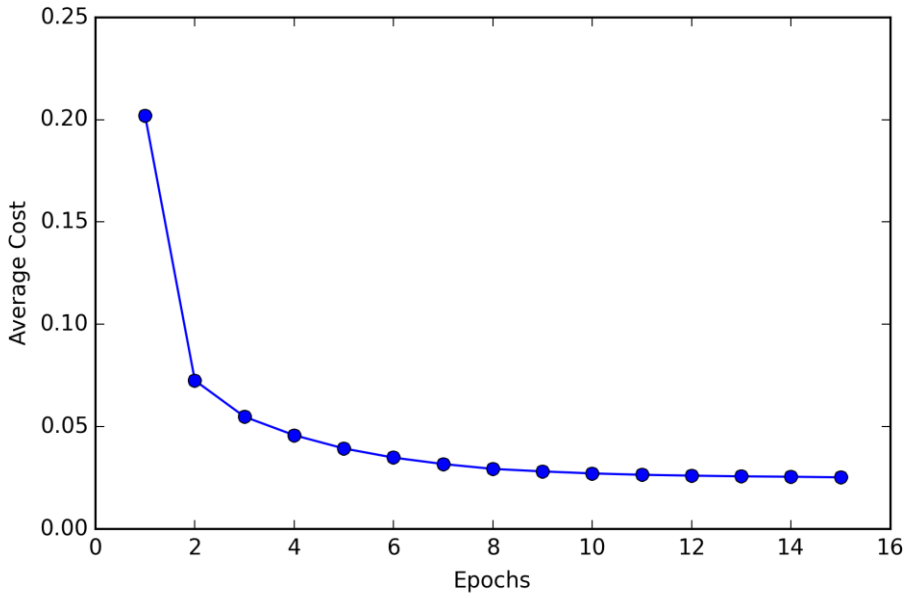


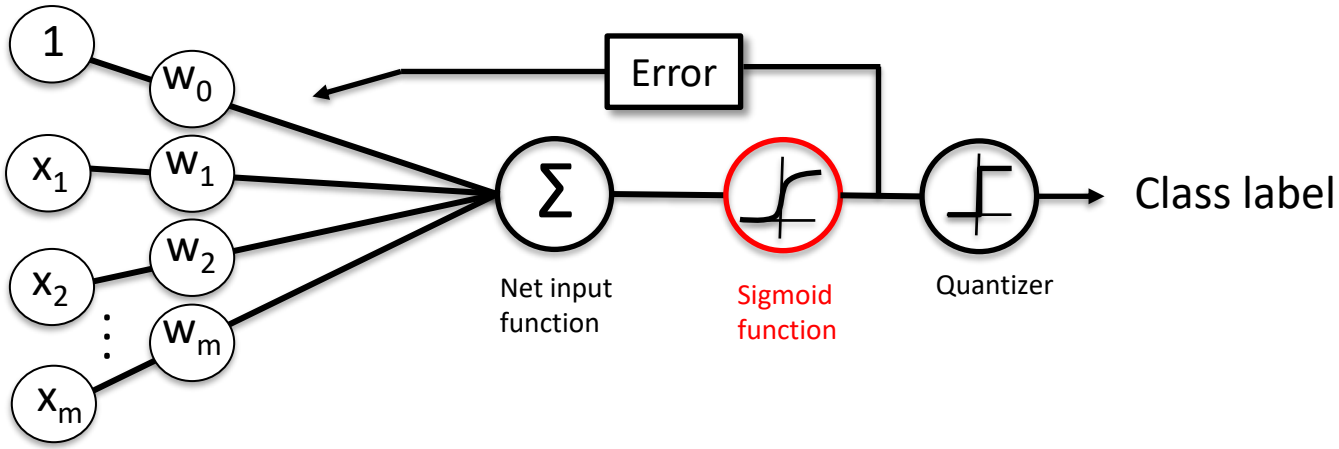


iris_adaline_sgd_cost_convergence_1.png

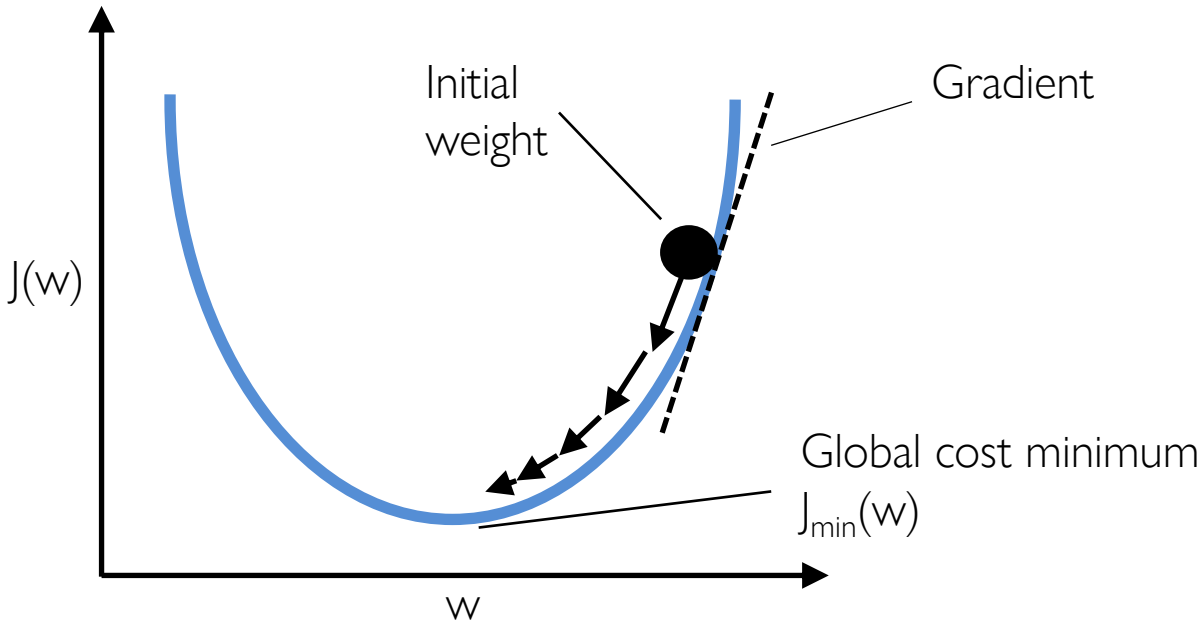
iris_adaline_sgd_cost_convergence_2.png

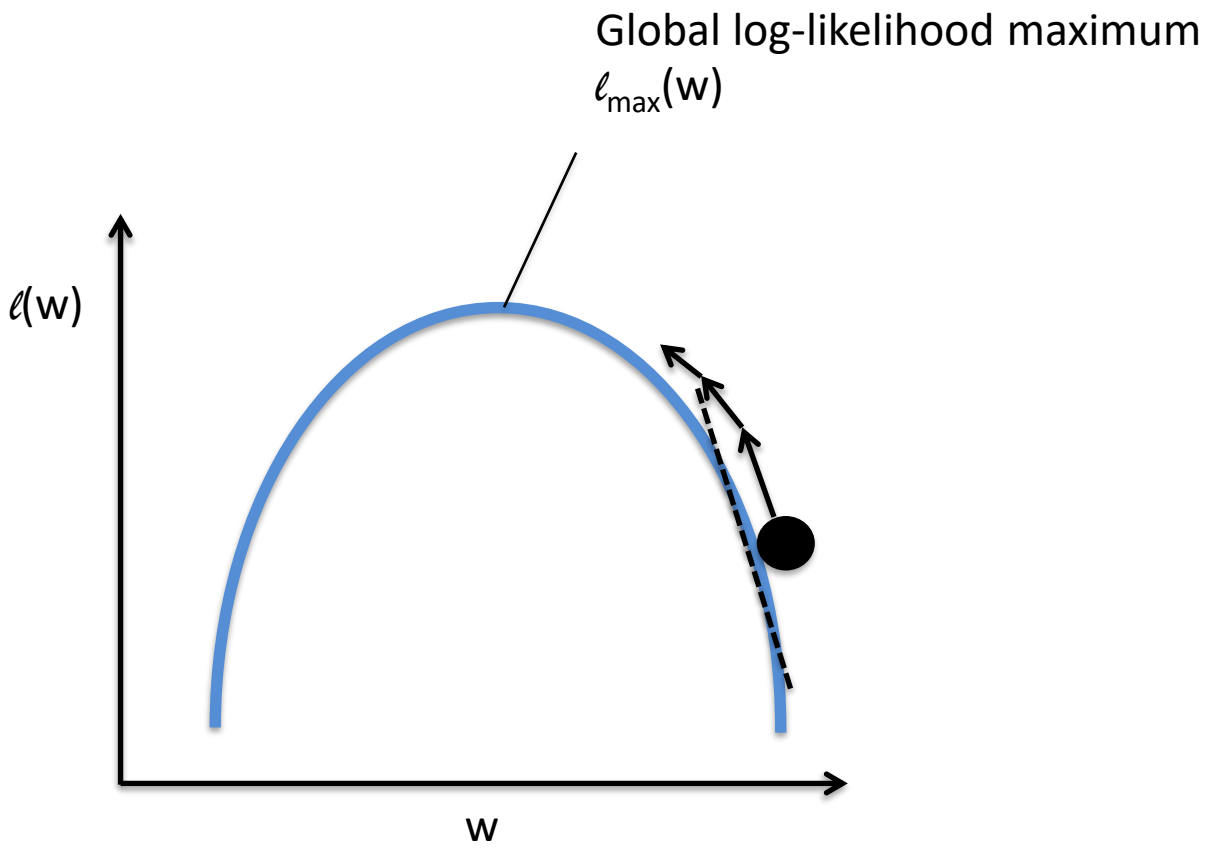


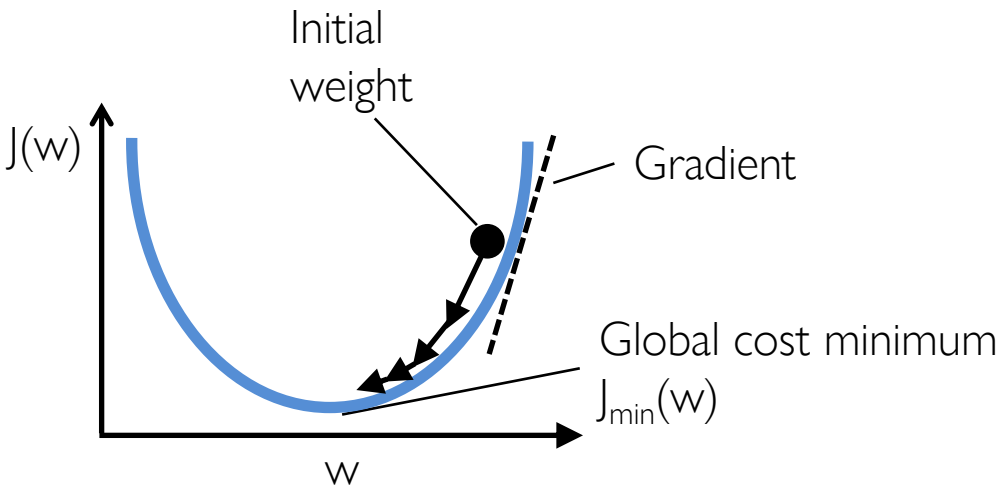


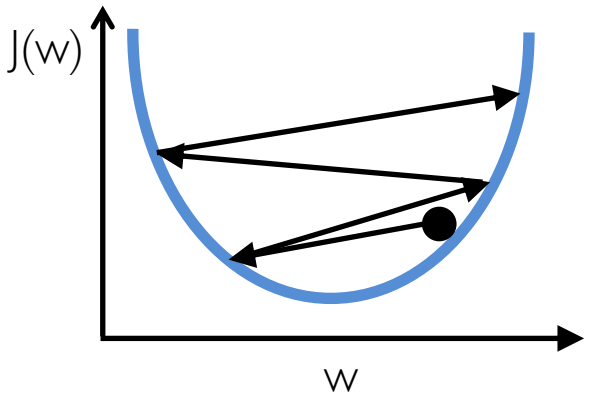

1
w_0
x_1
w_1
x_2
w_2
w_m
x_m
Error
Σ
Net input function
Sigmoid function
Quantizer
Class label

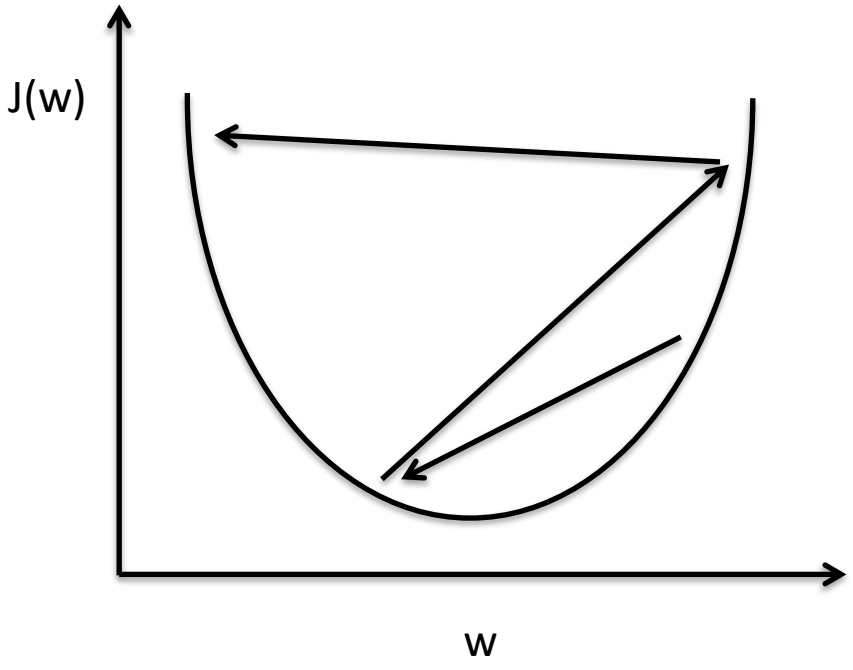
$$\begin{pmatrix} 1 & 2 & 3 \end{pmatrix} \times \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix} = 1\times4 + 2\times5 + 3\times6 = 32$$


Initial
weight
Gradient
J(w)
Global cost minimum
$J_{min}(w)$
w

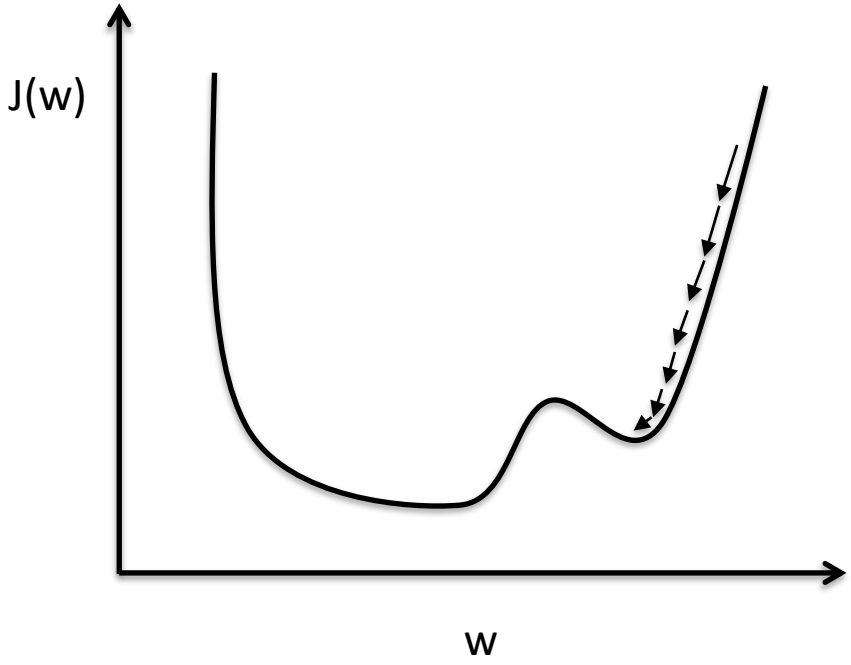

Global log-likelihood maximum
$\ell_{max}(w)$
$\ell(w)$
w


Initial
weight
J(w)
Gradient
Global cost minimum
$J_{min}(w)$
w

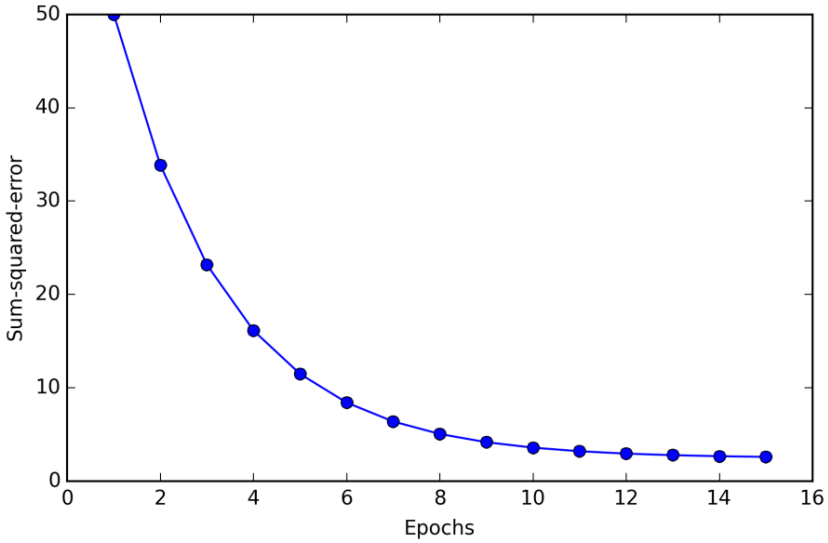

J(w)
w

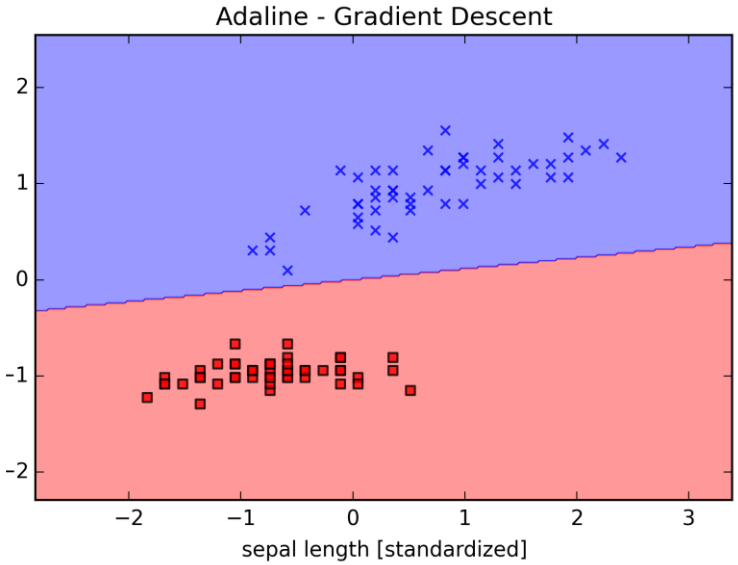


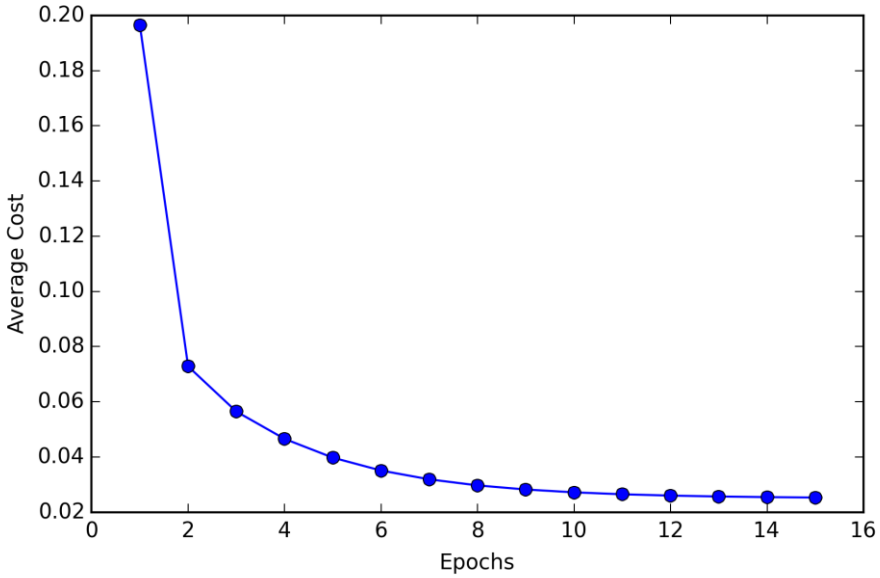
Large learning rate: Overshooting.

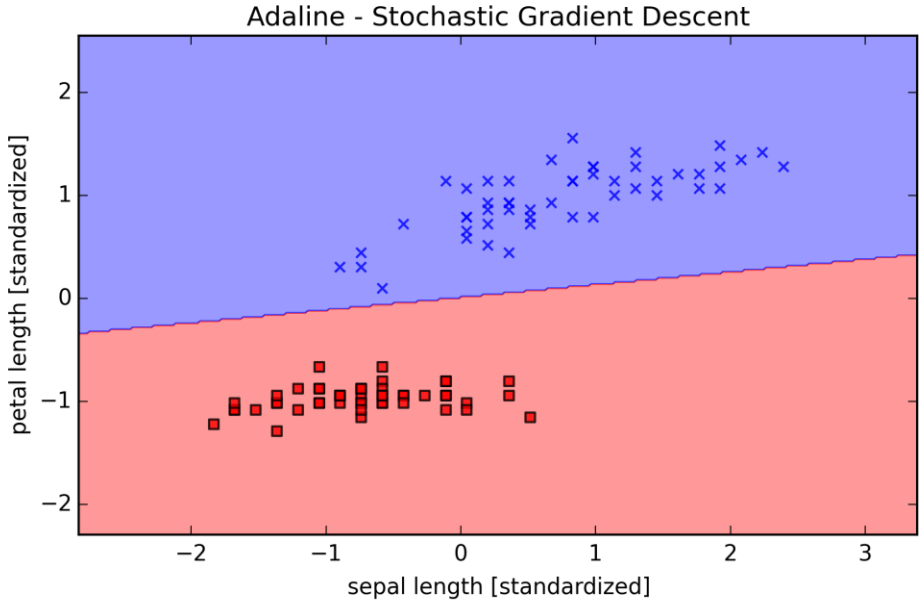


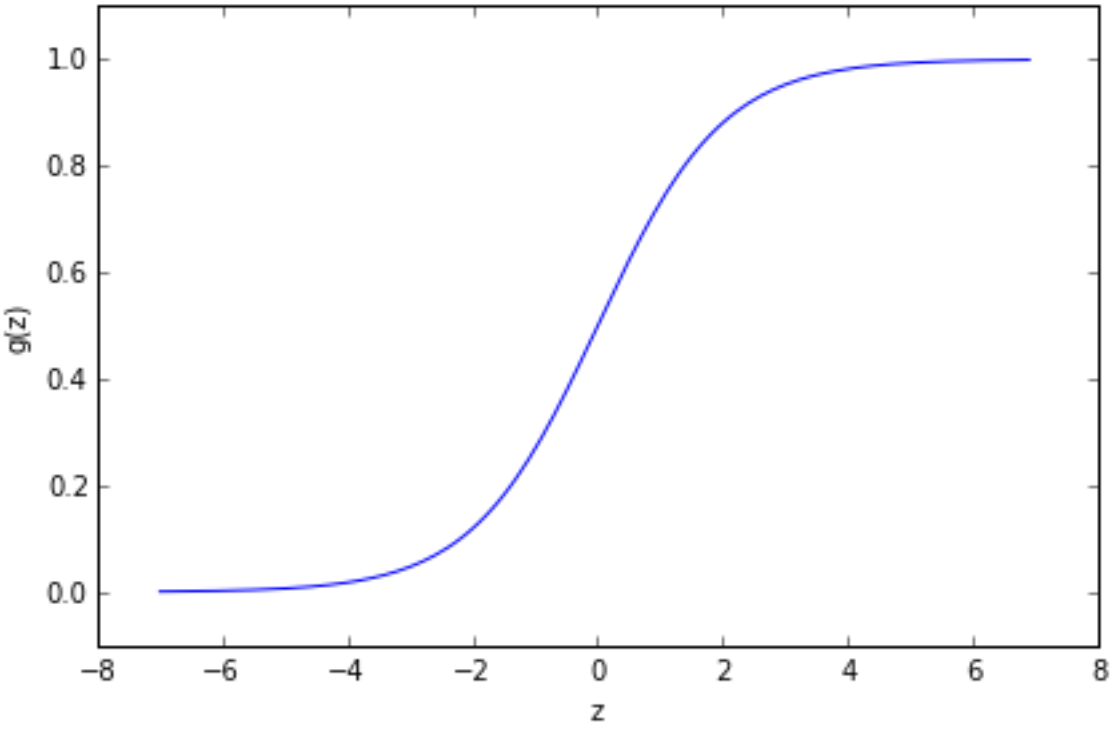
Small learning rate: Many iterations until convergence and trapping in local minima.

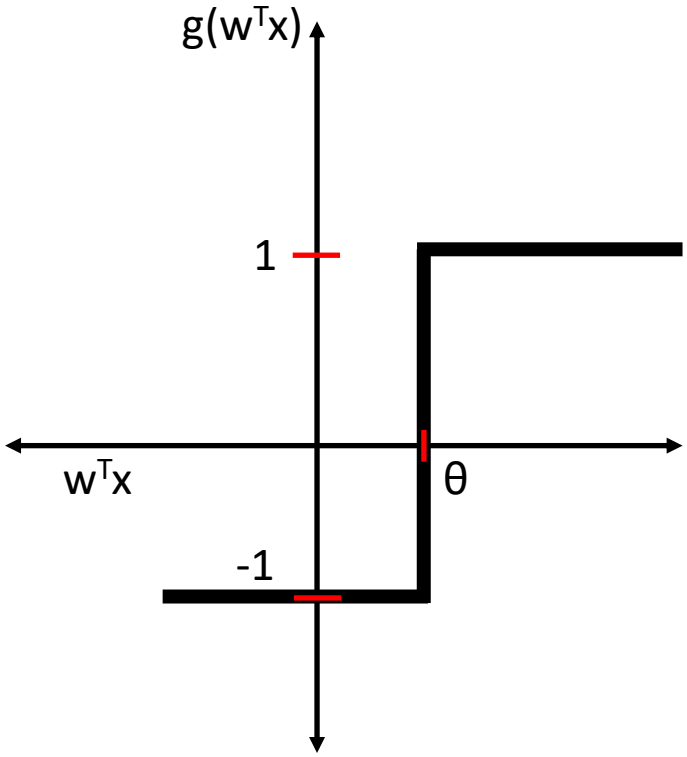

Sum-squared-error
Epochs
0
10
20
30
40
50
0
2
4
6
8
10
12
14
16


Adaline - Gradient Descent
sepal length [standardized]
−2
−1
0
1
2
3
−2
−1
0
1
2

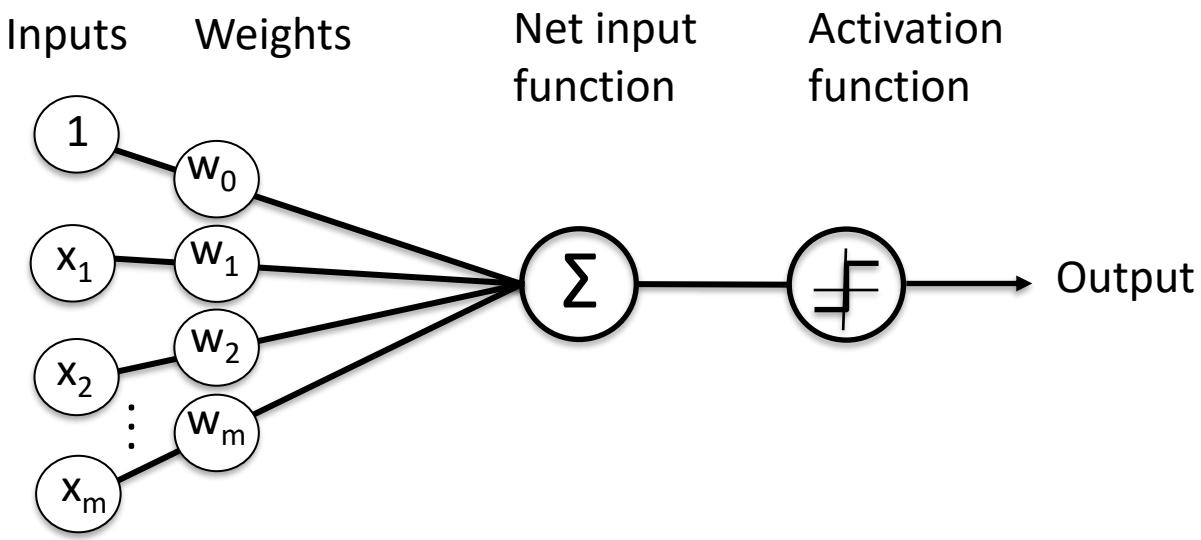

Average Cost
Epochs
0.20
0.18
0.16
0.14
0.12
0.10
0.08
0.06
0.04
0.02
0
2
4
6
8
10
12
14
16


Adaline - Stochastic Gradient Descent
petal length [standardized]
sepal length [standardized]
2
1
0
−1
−2
−2
−1
0
1
2
3

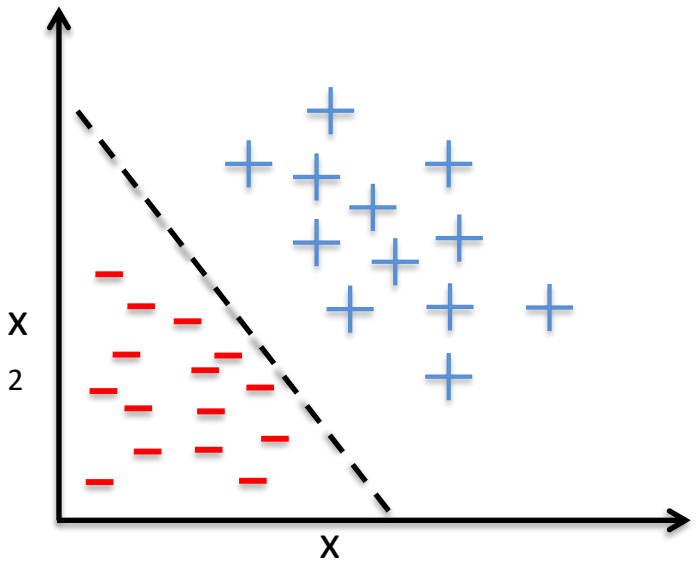

1.0
0.8
0.6
0.4
0.2
0.0
g(z)
−8
−6
−4
−2
0
2
4
6
8
z



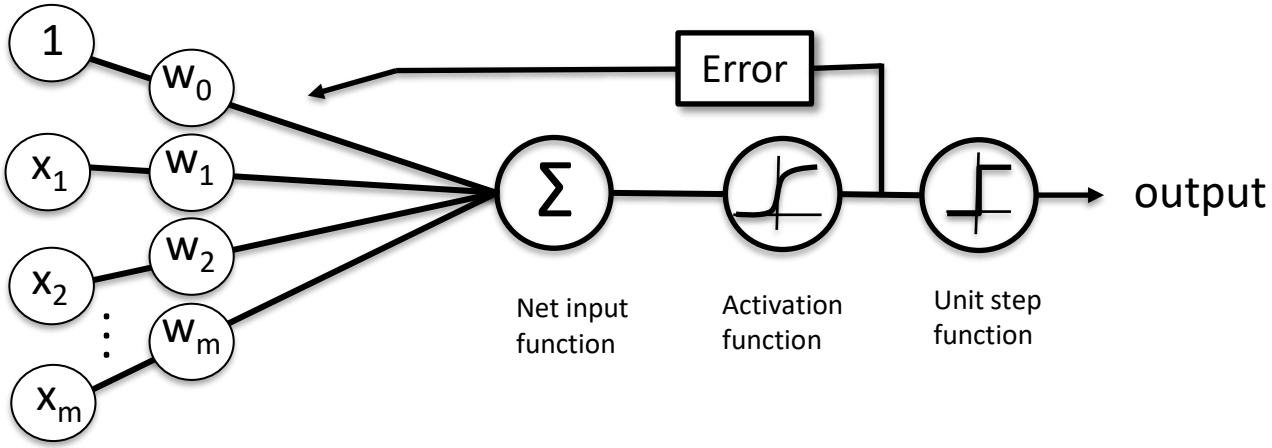
Unit step function.



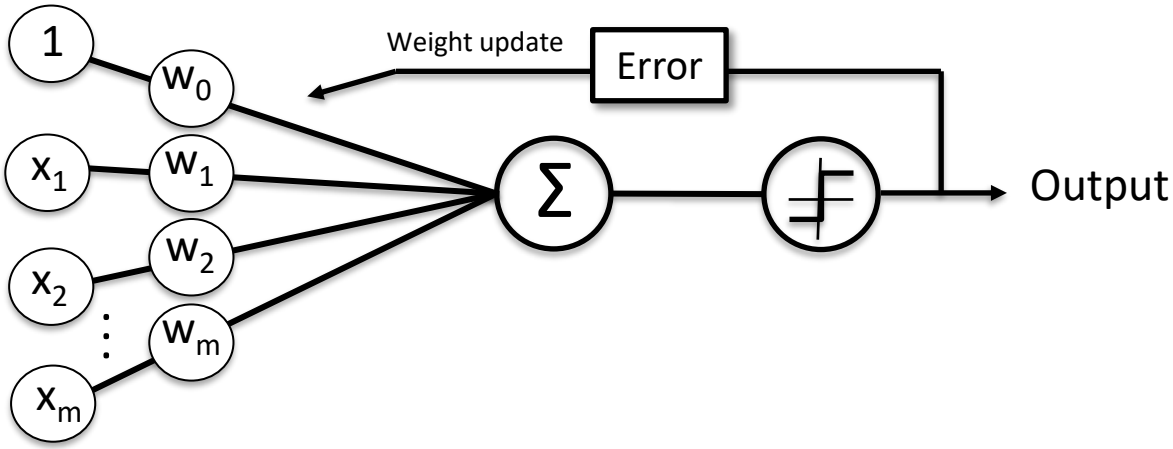
Schematic of Rosenblatt's perceptron.



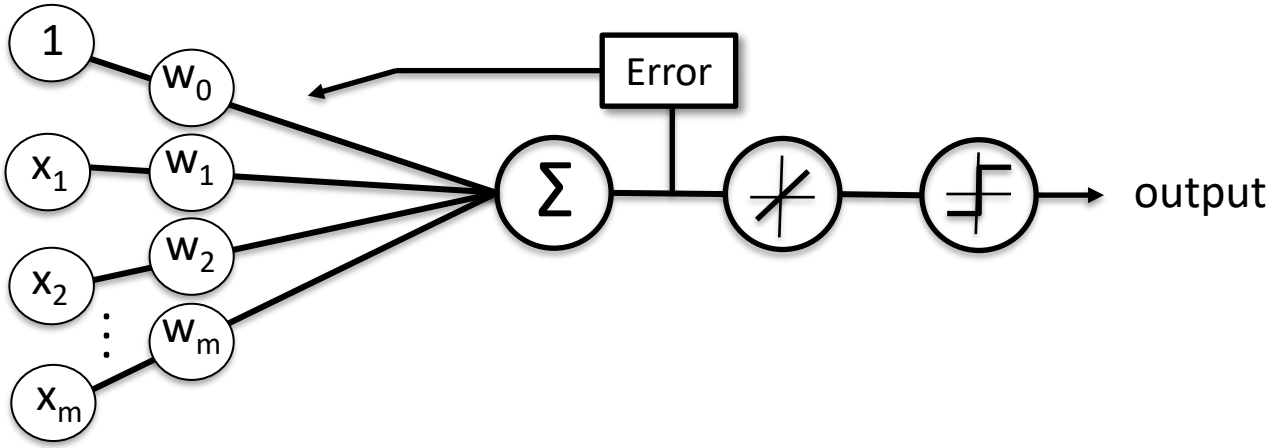
Example of a linear decision boundary for binary classification.



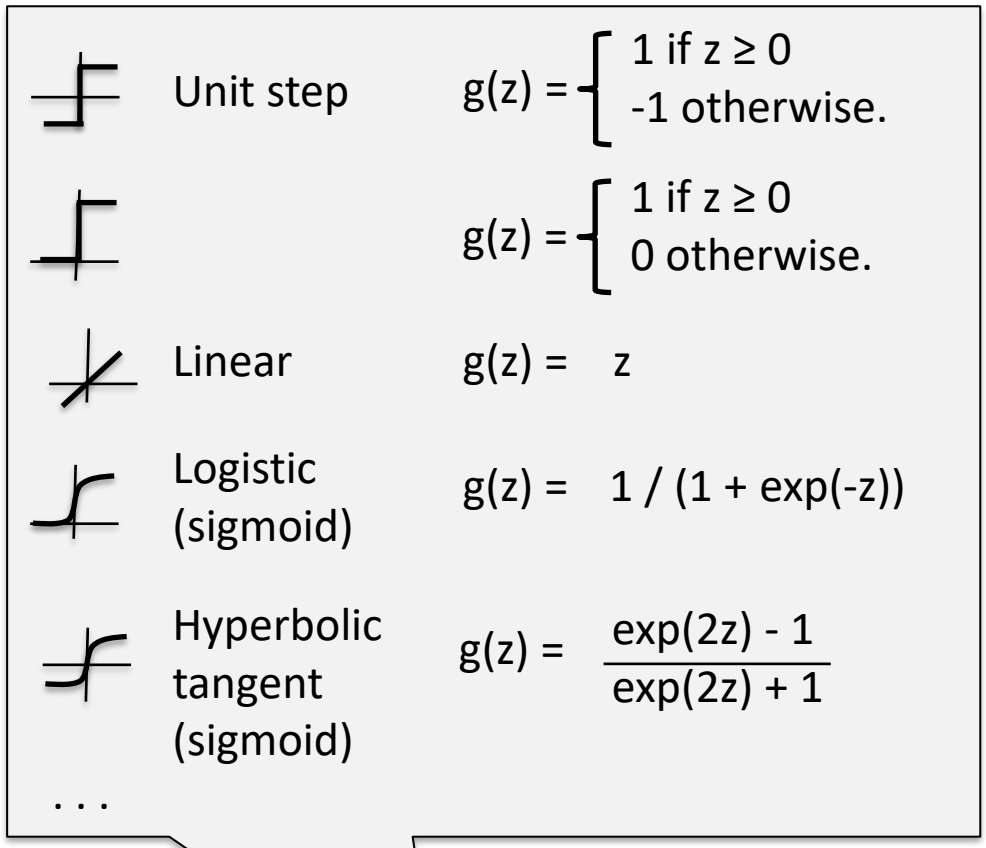
Schematic of a logistic regression classifier.

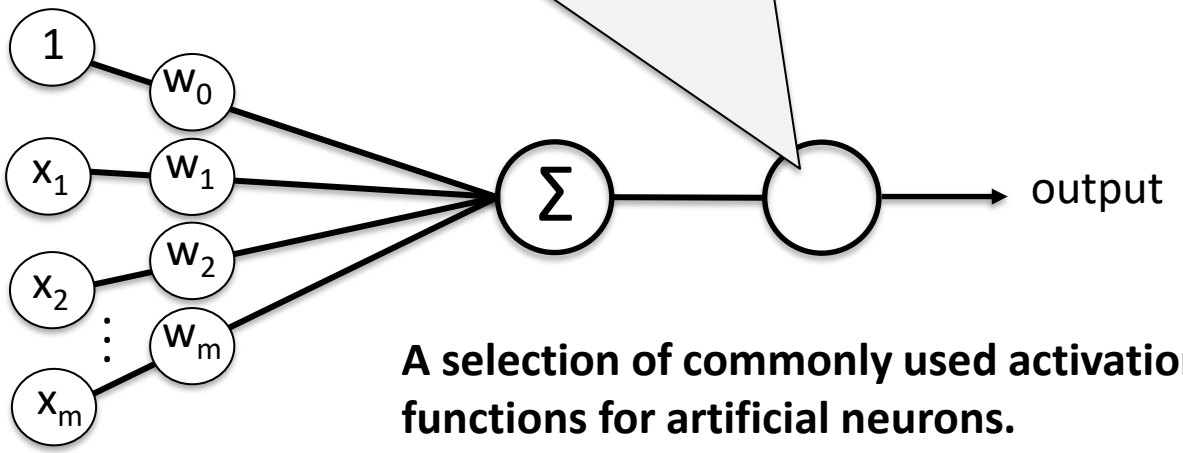


Schematic of a perceptron classifier.

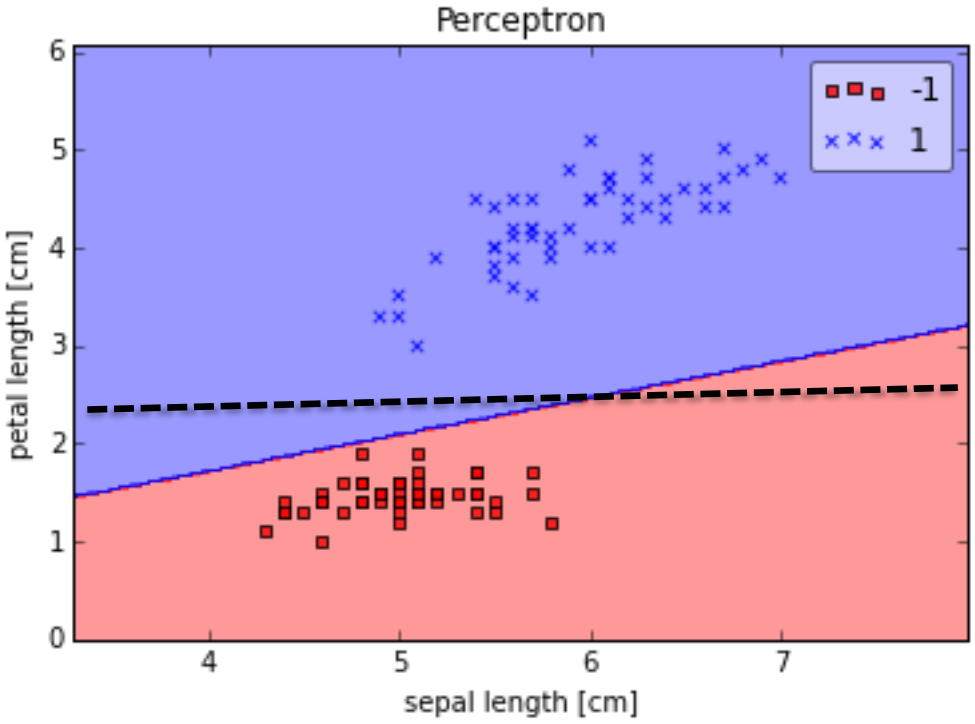


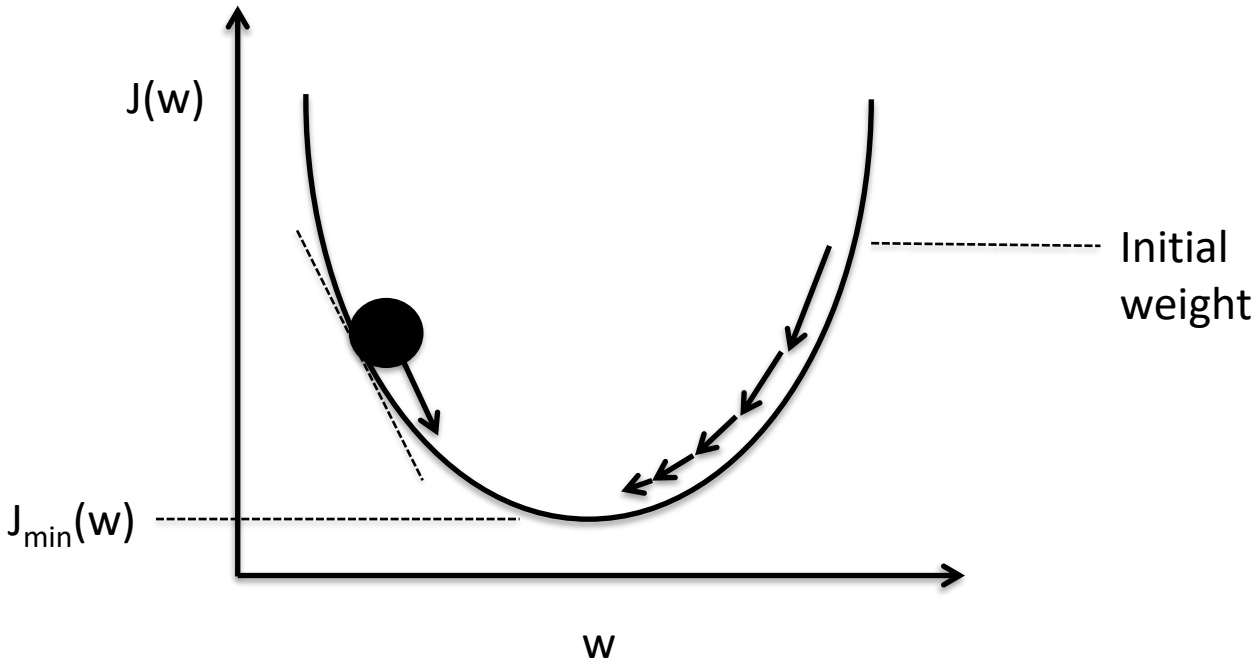
Schematic of an Adaline classifier.



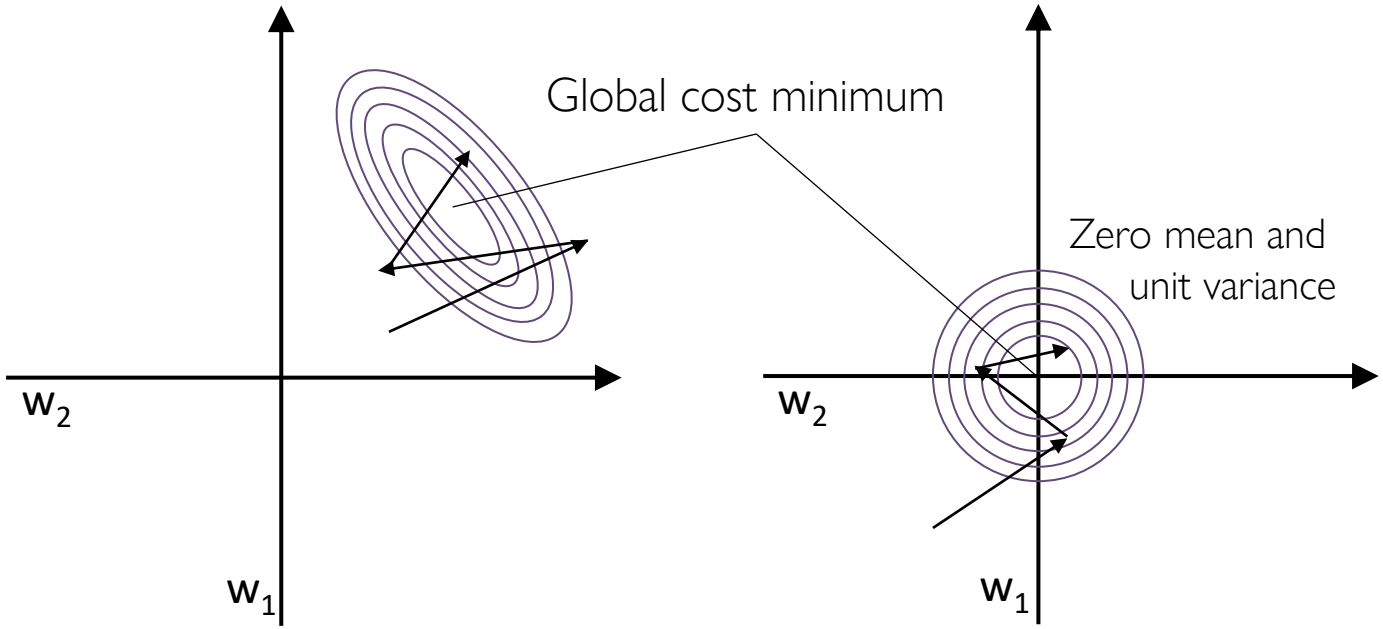


A selection of commonly used activation functions for artificial neurons.


Perceptron
-1
1
petal length [cm]
sepal length [cm]
0
1
2
3
4
5
6
4
5
6
7



Schematic of gradient descent.


Global cost minimum
Zero mean and
unit variance
w2
w1
w2
w1