An Introduction to Deep Learning with TensorFlow

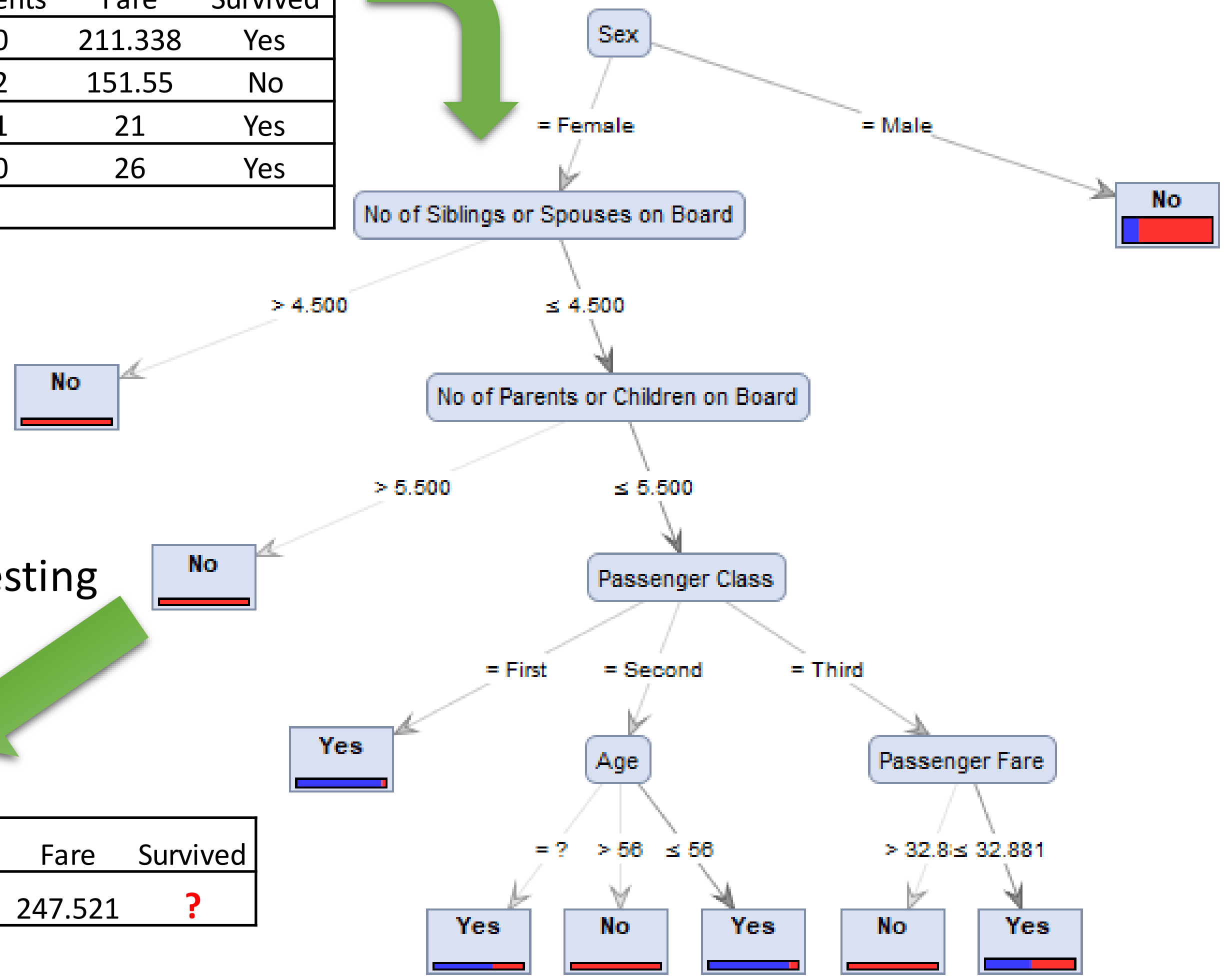
Steven H. H. Ding

School of Information Studies

Last Workshop - Classification with Decision Tree

Passenger Class	Gender	age	Siblings	Parents	Fare	Survived
First	Female	29	0	0	211.338	Yes
First	Female	2	1	2	151.55	No
Second	Female	21	0	1	21	Yes
Second	Female	19	0	0	26	Yes
...						

Training



Testing

Passenger Class	Gender	age	Siblings	Parents	Fare	Survived
First	Male	24	0	1	247.521	?

So, what is Deep Learning

So, what is Deep Learning

So, what is Deep Learning

So, what is Deep Learning

So, what is Deep Learning

So, what is Deep Learning

So, what is Deep Learning

So, what is Deep Learning

Non-linearity

So, what is Deep Learning

Non-linearity

Feature Learning

So, what is Deep Learning

Non-linearity

Auto Encoder

Feature Learning

So, what is Deep Learning

Non-linearity

Recursive Network

Auto Encoder

Feature Learning

So, what is Deep Learning

Non-linearity

Recursive Network

Deep Brief Network

Auto Encoder

Feature Learning

So, what is Deep Learning

Non-linearity

Recursive Network

Deep Brief Network

Stochastic Gradient Descend

Auto Encoder

Feature Learning

So, what is Deep Learning

Non-linearity

Recursive Network

Deep Brief Network

Stochastic Gradient Descend

Convolutionary Neural Network

Auto Encoder

Feature Learning

So, what is Deep Learning

Non-linearity

Recursive Network

Deep Brief Network

Back Propagation

Stochastic Gradient Descend

Convolutionary Neural Network

Auto Encoder

Feature Learning

So, what is Deep Learning

Non-linearity

Recursive Network

Deep Brief Network

Back Propagation

Recurrent Network

Stochastic Gradient Descend

Convolutionary Neural Network

Auto Encoder

Feature Learning

So, what is Deep Learning

Non-linearity

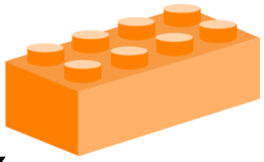
‘Deep Learning’ means using a ***neural network*** with several layers of nodes between input and output.

It's deep if it has more than one stage of non-linear feature transformation.

Agenda

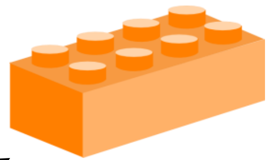
- Logistic Regression
- Neural Network
- Convolutional Neural Network
- Why deep learning works
- Hands-on example using TensorFlow
- Q&A

Agenda



- Logistic Regression
- Neural Network
- Convolutional Neural Network
- Why deep learning works
- Hands-on example using TensorFlow
- Q&A

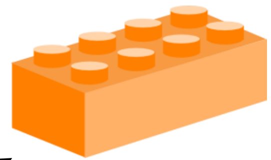
Agenda



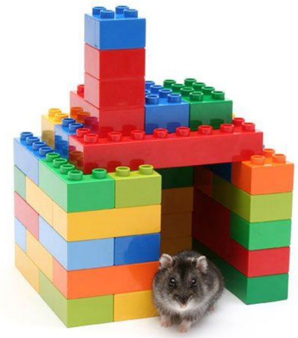
- Logistic Regression

- Neural Network
- Convolutional Neural Network
- Why deep learning works
- Hands-on example using TensorFlow
- Q&A

Agenda



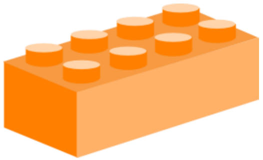
- Logistic Regression



- Neural Network
- Convolutional Neural Network

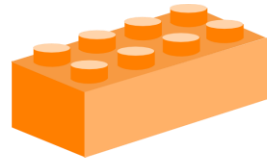
- Why deep learning works
- Hands-on example using TensorFlow
- Q&A

Logistic Regression

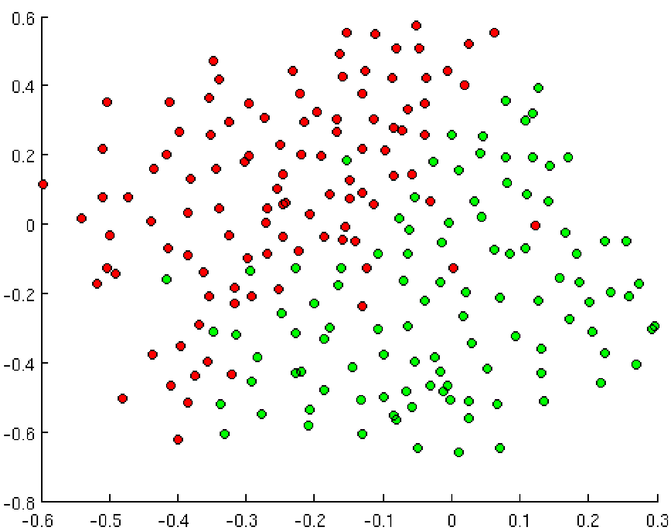


Passenger Class	Age	Survived
1	29	1
1	2	0
2	21	1
2	19	1
...		

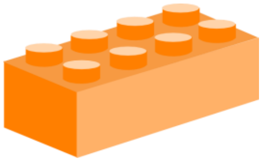
Logistic Regression



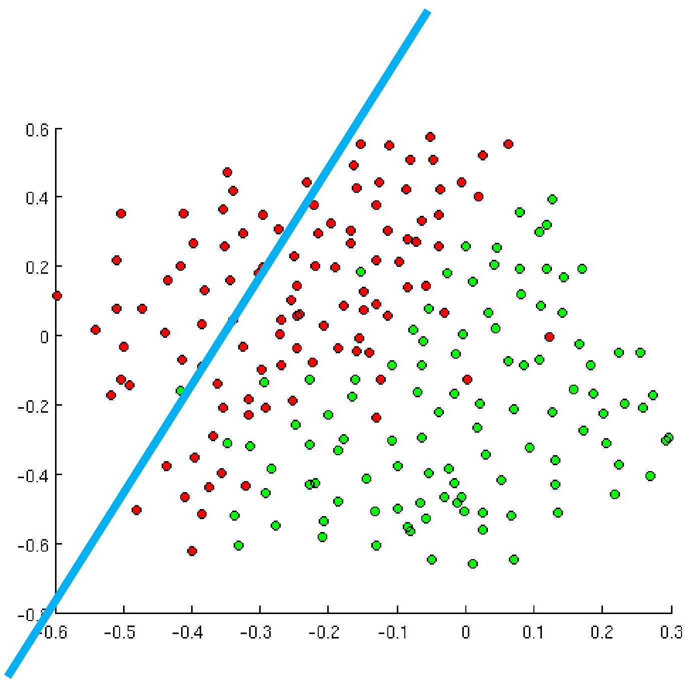
Passenger Class	Age	Survived
1	29	1
1	2	0
2	21	1
2	19	1
...		



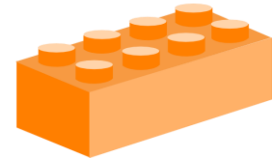
Logistic Regression



Passenger Class	Age	Survived
1	29	1
1	2	0
2	21	1
2	19	1
...		

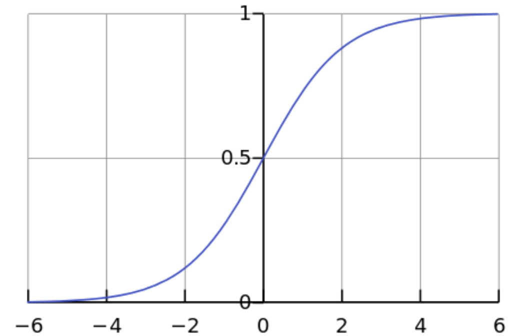


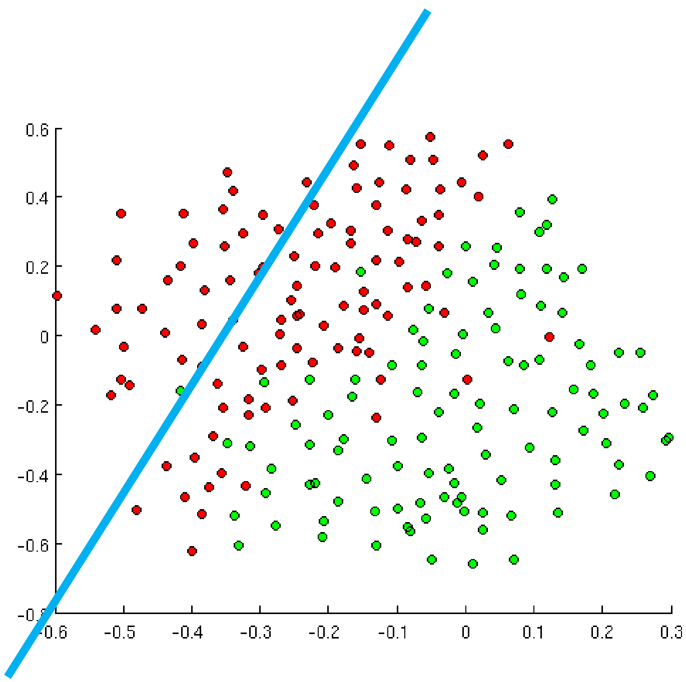
Logistic Regression

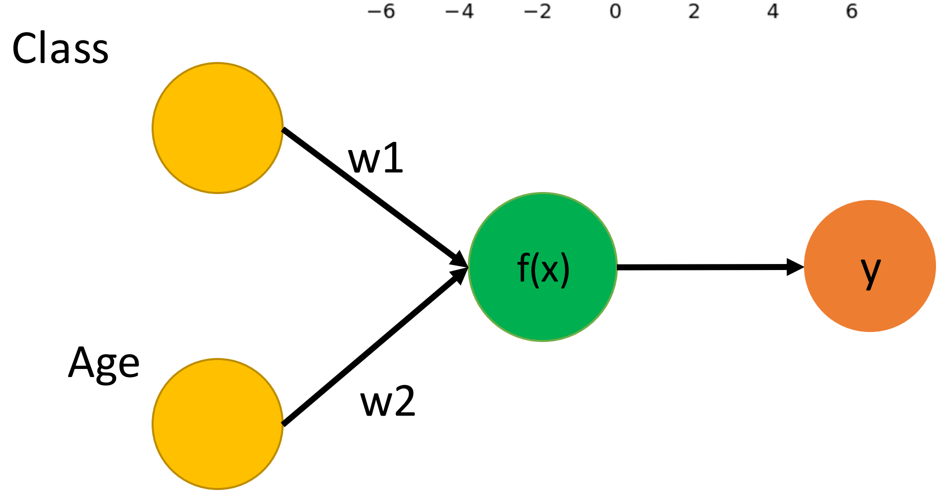


Passenger Class	Age	Survived
1	29	1
1	2	0
2	21	1
2	19	1
...		

$$f(x) = \frac{1}{1 + e^{-x}}$$



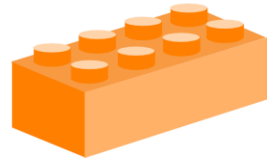




a = w1 * Class + w2 * Age

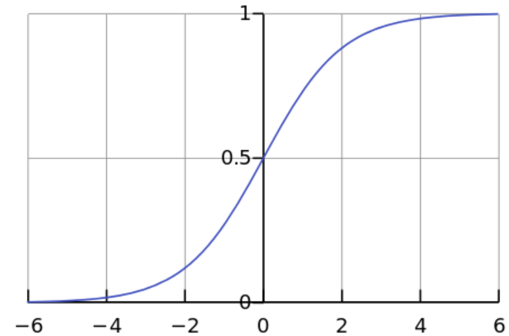
y = f(a)

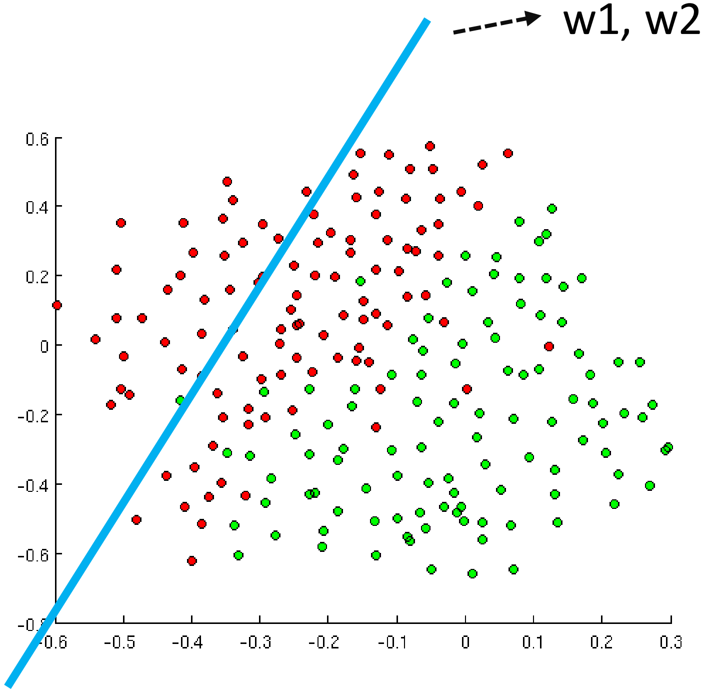
Logistic Regression

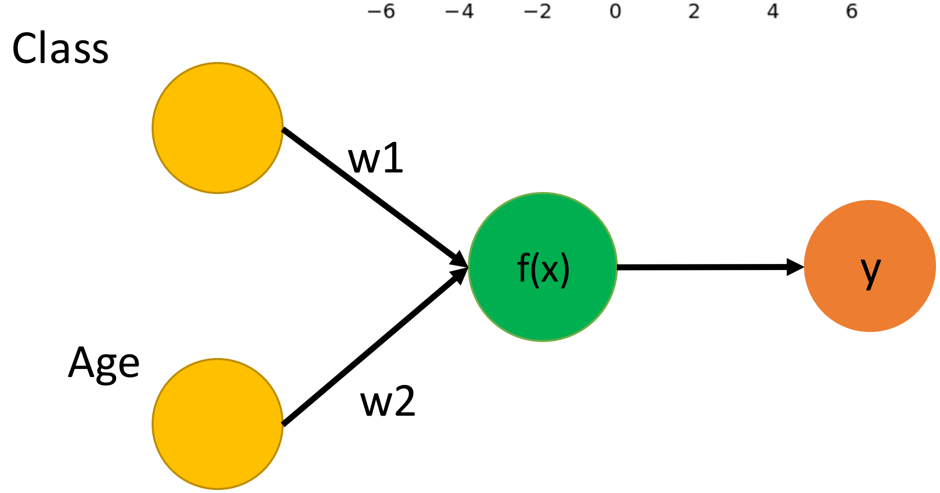


Passenger Class	Age	Survived
1	29	1
1	2	0
2	21	1
2	19	1
...		

$$f(x) = \frac{1}{1+e^{-x}}$$



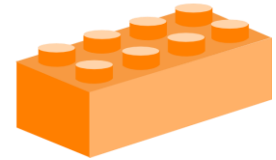




a = w1 * Class + w2 * Age

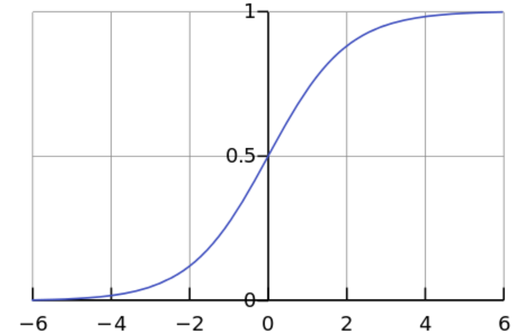
y = f(a)

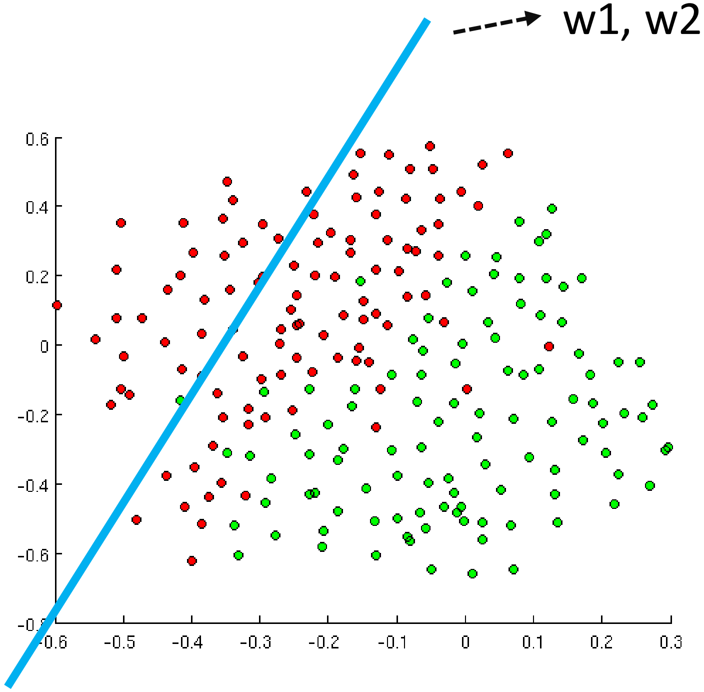
Logistic Regression

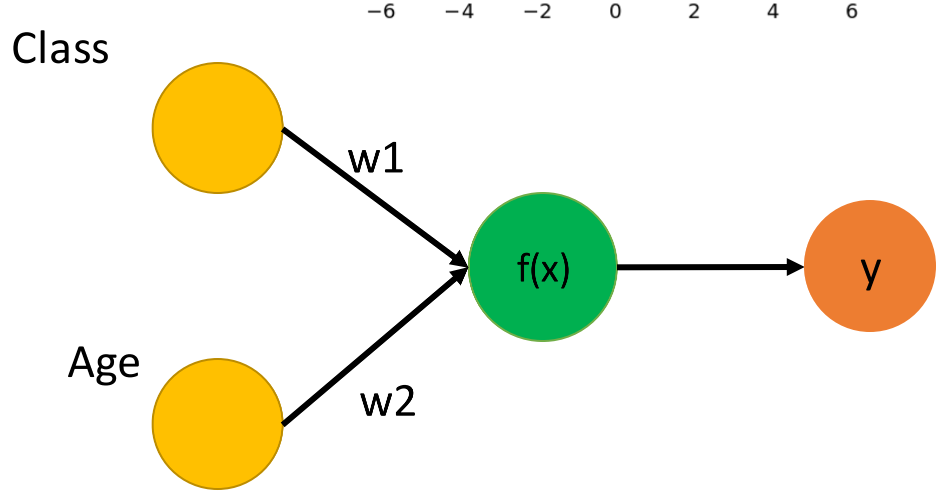


Passenger Class	Age	Survived
1	29	1
1	2	0
2	21	1
2	19	1
...		

$$f(x) = \frac{1}{1 + e^{-x}}$$



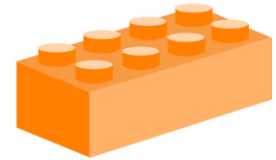




a = w1 * Class + w2 * Age

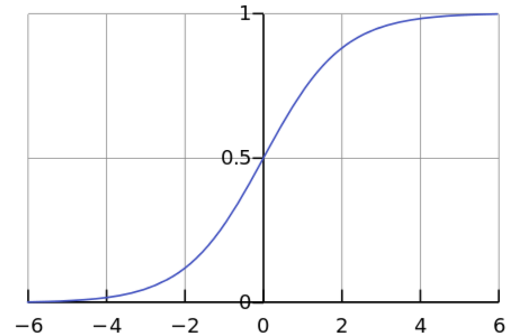
y = f(a)

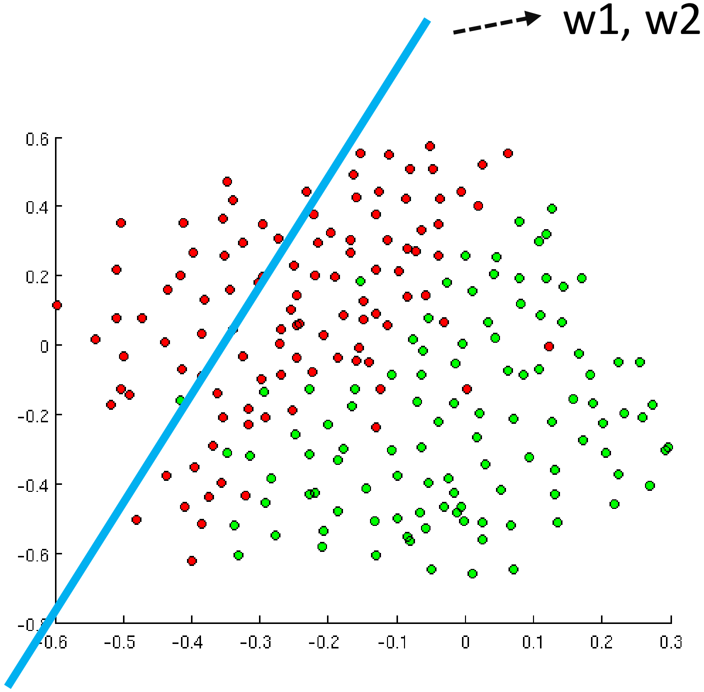
Logistic Regression

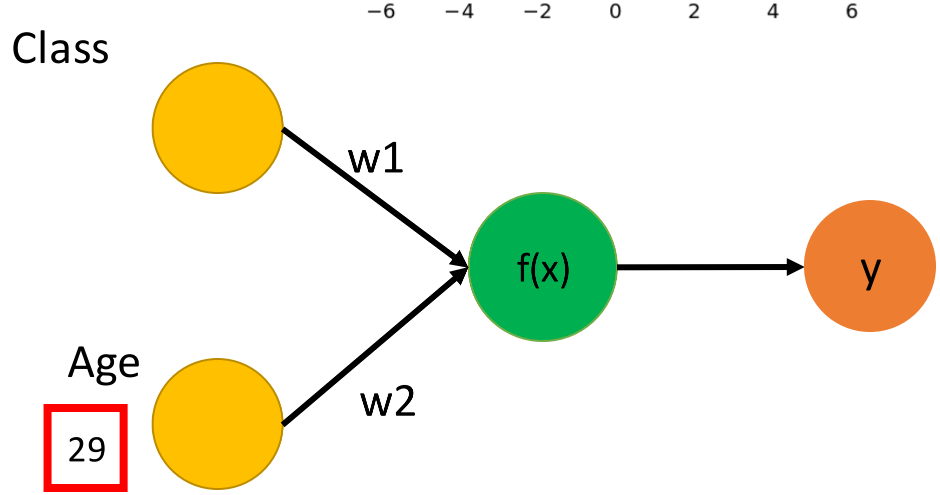


Passenger Class	Age	Survived
1	29	1
1	2	0
2	21	1
2	19	1
...		

$$f(x) = \frac{1}{1+e^{-x}}$$



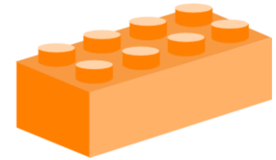




a = w1 * Class + w2 * Age

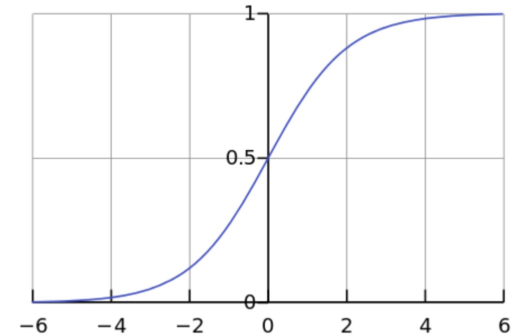
y = f(a)

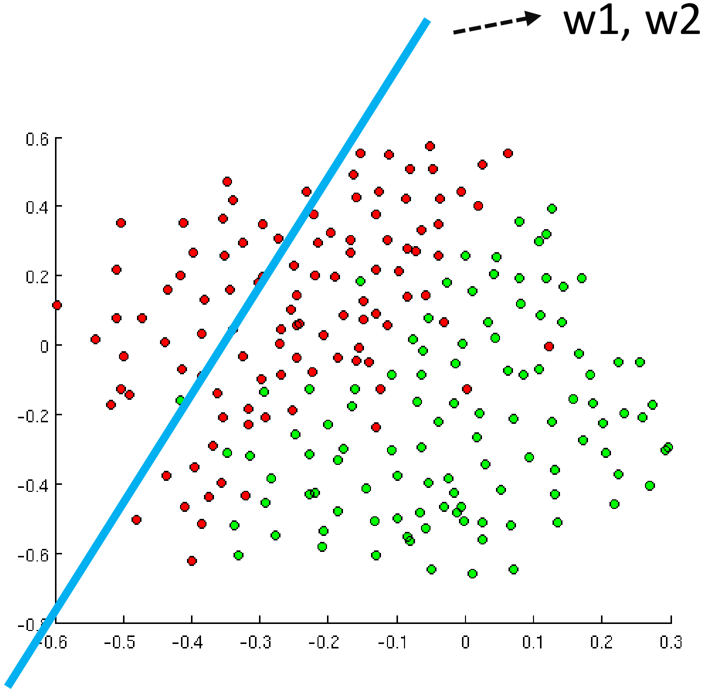
Logistic Regression

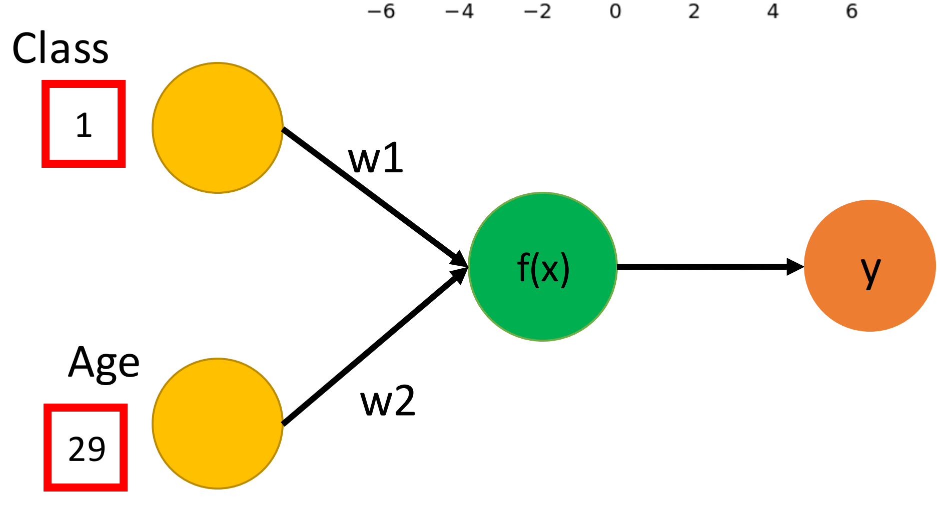


Passenger Class	Age	Survived
1	29	1
1	2	0
2	21	1
2	19	1
...		

$$f(x) = \frac{1}{1 + e^{-x}}$$



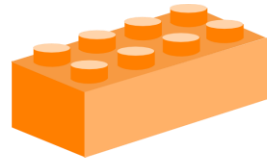




a = w1 * Class + w2 * Age

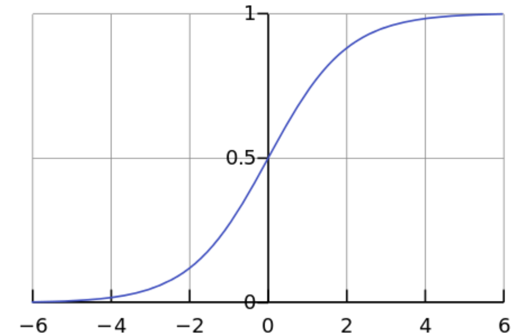
y = f(a)

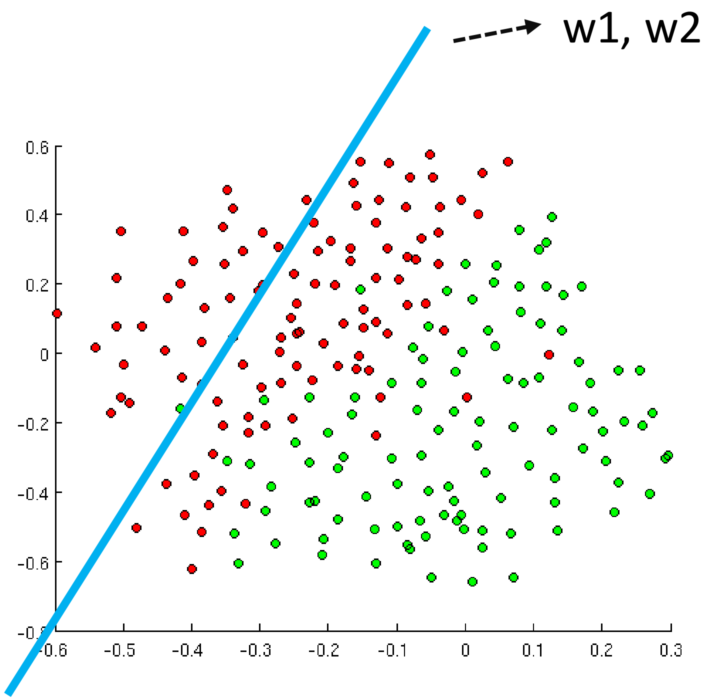
Logistic Regression

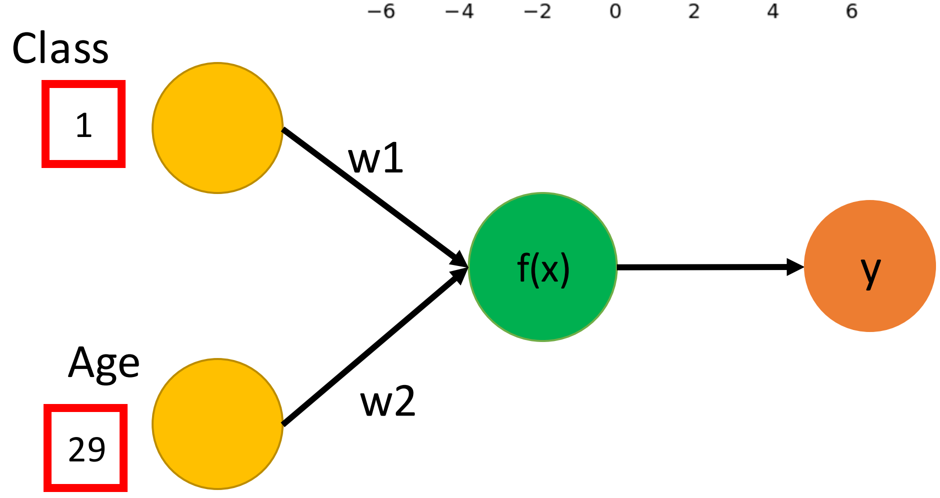


Passenger Class	Age	Survived
1	29	1
1	2	0
2	21	1
2	19	1
…		

$$f(x) = \frac{1}{1 + e^{-x}}$$





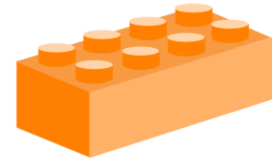


a = w1 * Class + w2 * Age

a' = -2 * 1 + 0.1 * 29 = 0.9

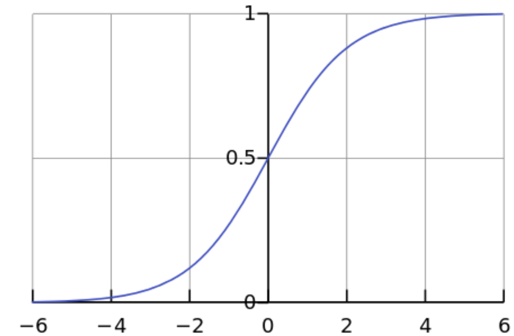
y = f(a)

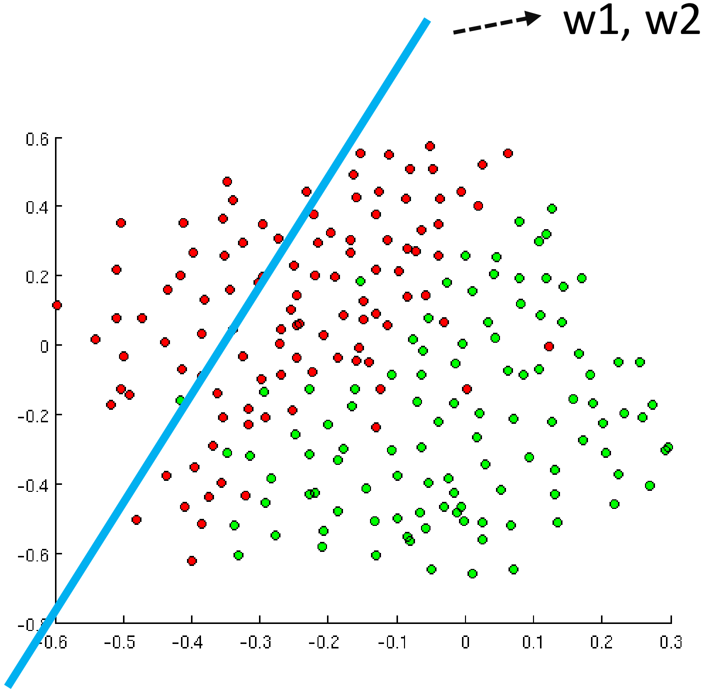
Logistic Regression

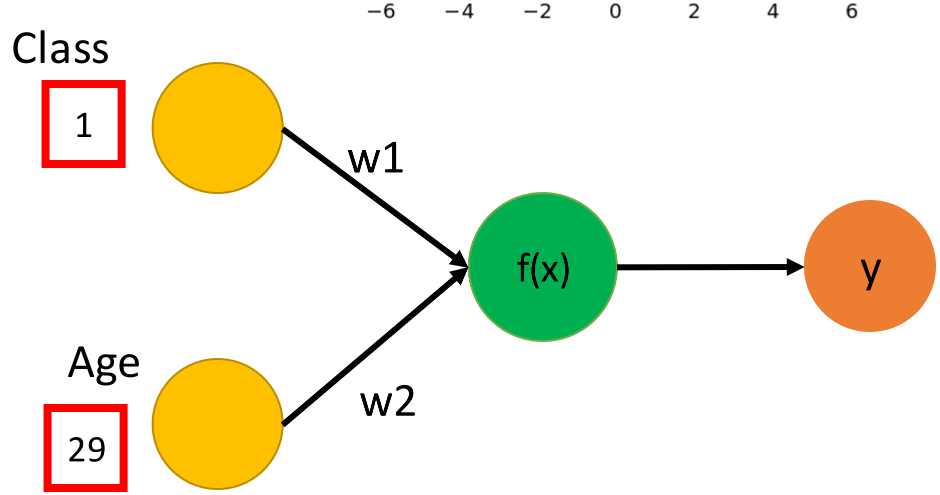


Passenger Class	Age	Survived
1	29	1
1	2	0
2	21	1
2	19	1
...		

$$f(x) = \frac{1}{1+e^{-x}}$$







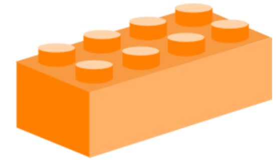
a = w1 * Class + w2 * Age

a' = -2 * 1 + 0.1 * 29 = 0.9

y = f(a)

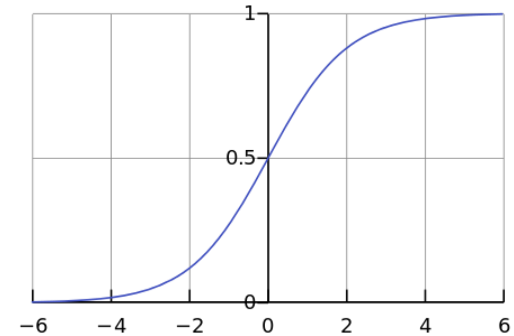
y' = f(0.9) = 0.71

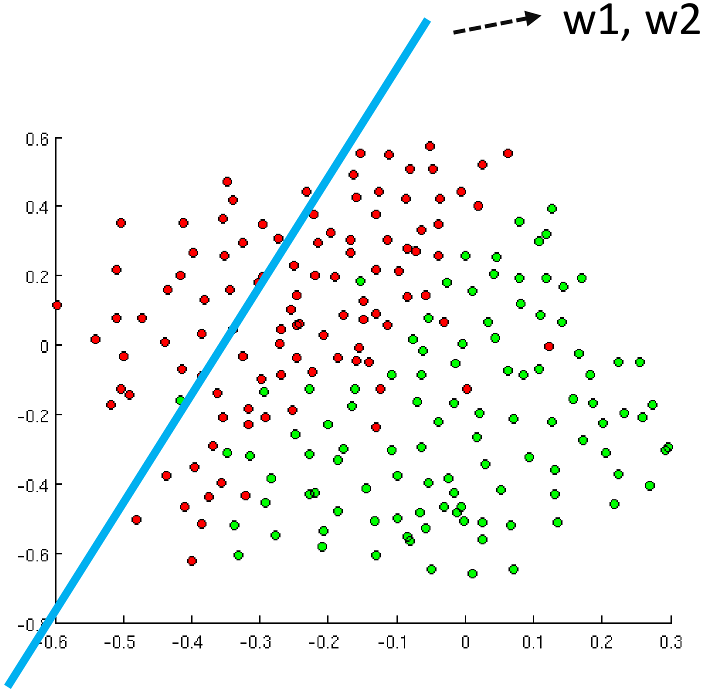
Logistic Regression

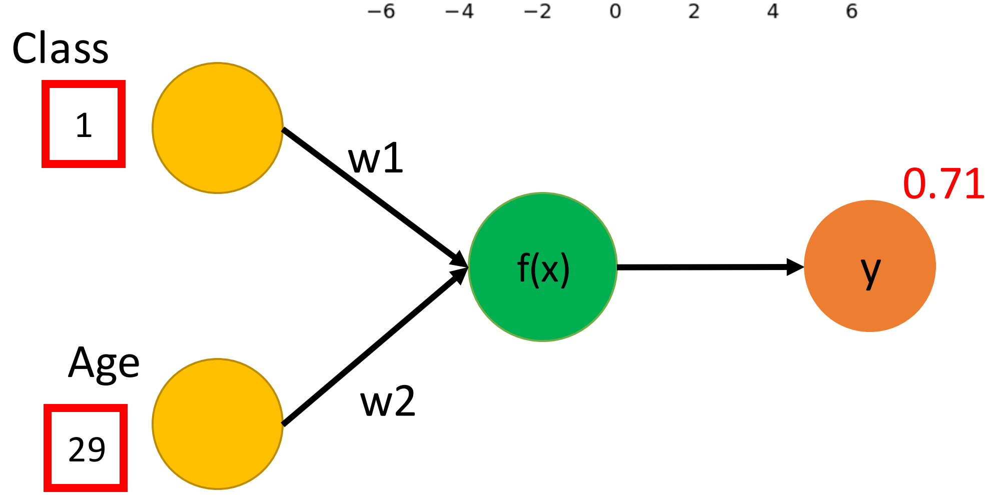


Passenger Class	Age	Survived
1	29	1
1	2	0
2	21	1
2	19	1
...		

$$f(x) = \frac{1}{1 + e^{-x}}$$







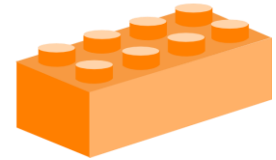
a = w1 * Class + w2 * Age

a' = -2 * 1 + 0.1 * 29 = 0.9

y = f(a)

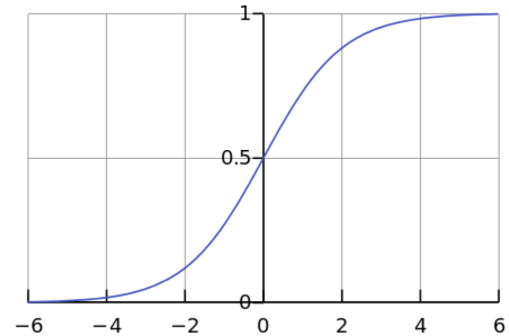
y' = f(0.9) = 0.71

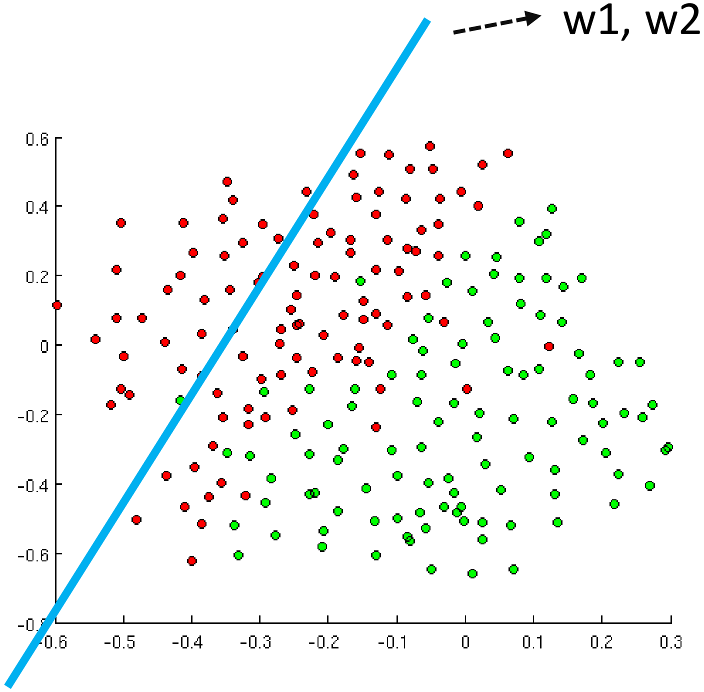
Logistic Regression

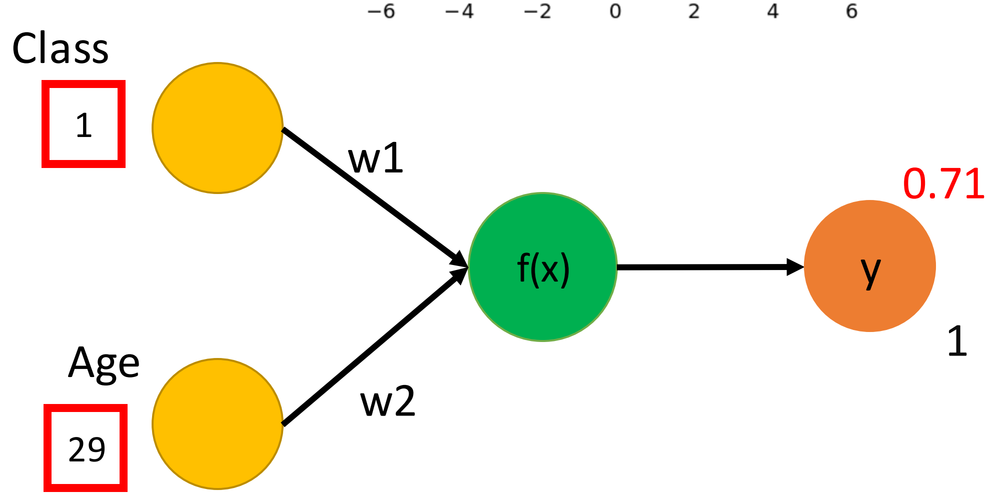


Passenger Class	Age	Survived
1	29	1
1	2	0
2	21	1
2	19	1
...		

$$f(x) = \frac{1}{1 + e^{-x}}$$







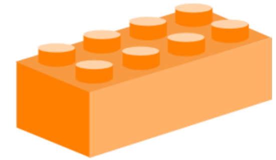
a = w1 * Class + w2 * Age

a' = -2 * 1 + 0.1 * 29 = 0.9

y = f(a)

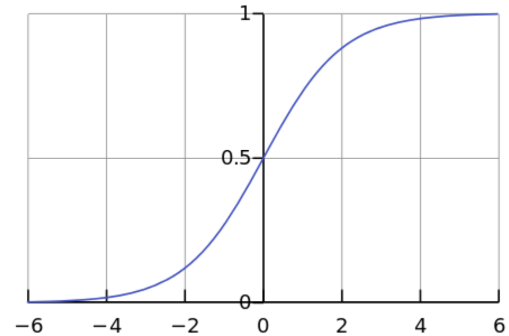
y' = f(0.9) = 0.71

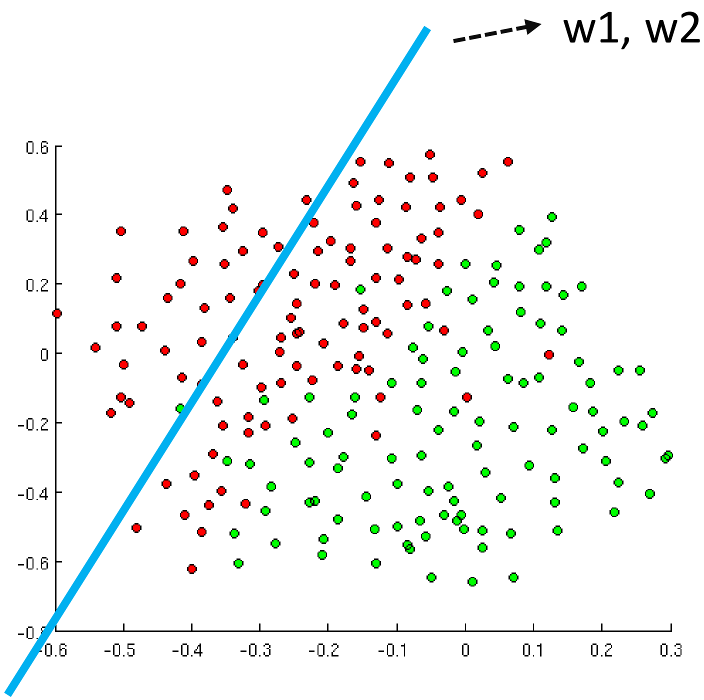
Logistic Regression

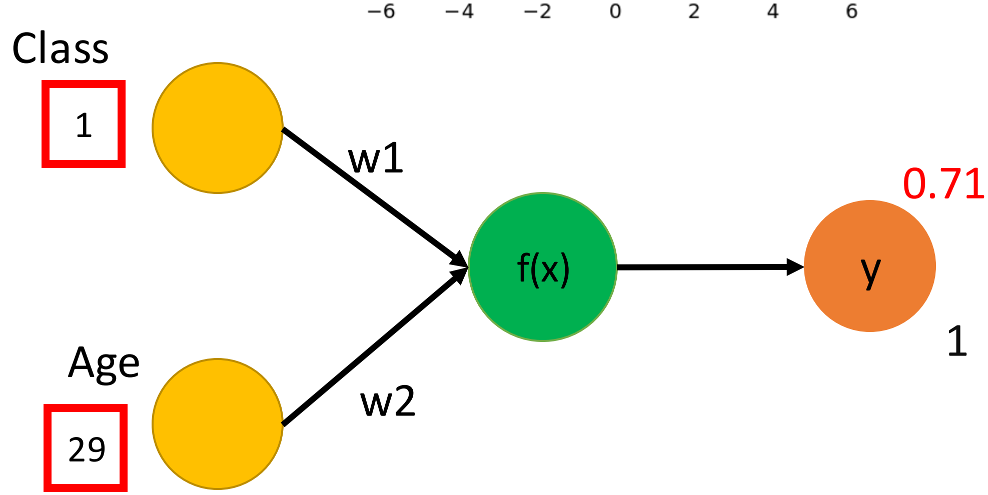


Passenger Class	Age	Survived
1	29	1
1	2	0
2	21	1
2	19	1
...		

$$f(x) = \frac{1}{1 + e^{-x}}$$







a = w1 * Class + w2 * Age

a' = -2 * 1 + 0.1 * 29 = 0.9

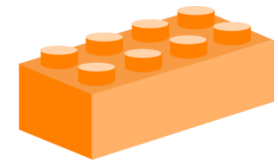
y = f(a)

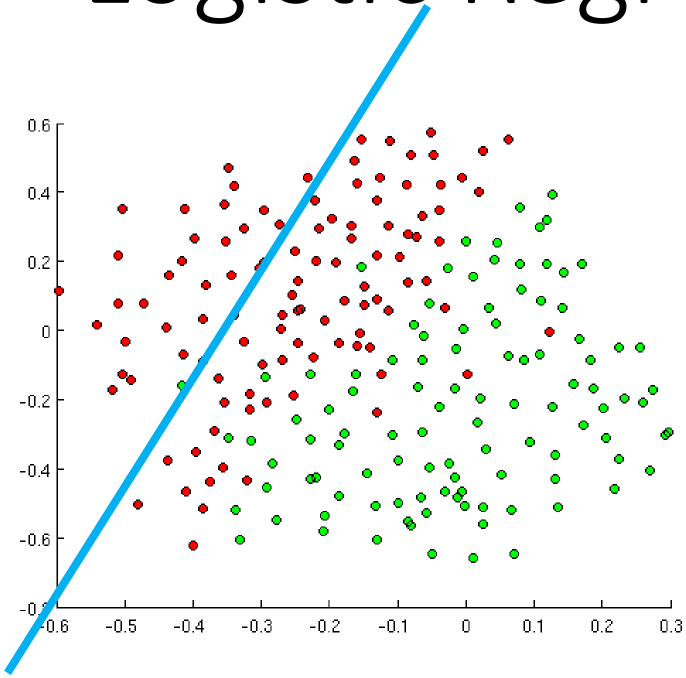
y' = f(0.9) = 0.71

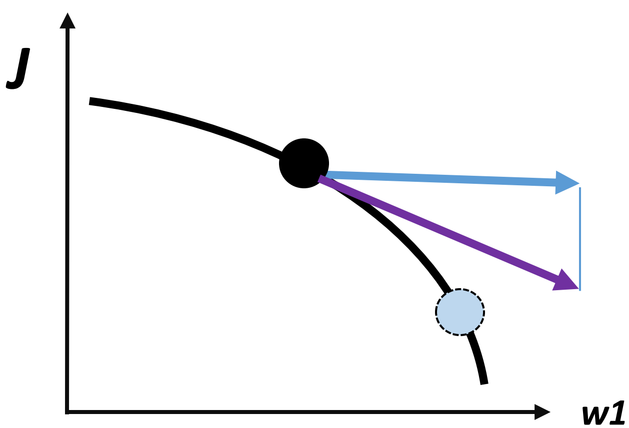
Total cost:

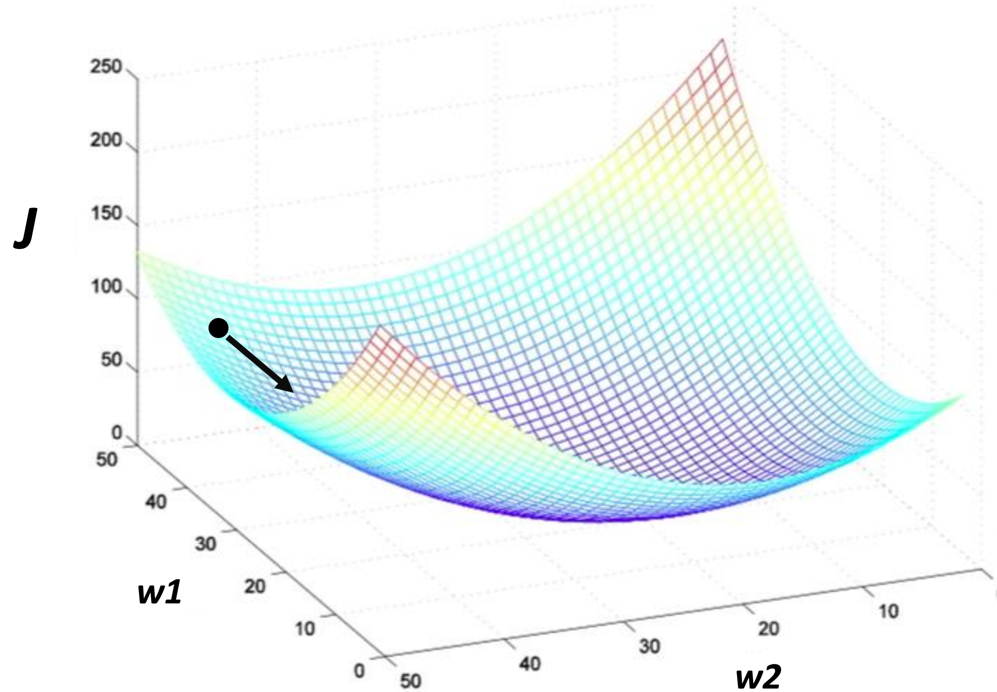
$$J(w) = \Sigma\,(y' - y)^2$$

Logistic Regression cont'd





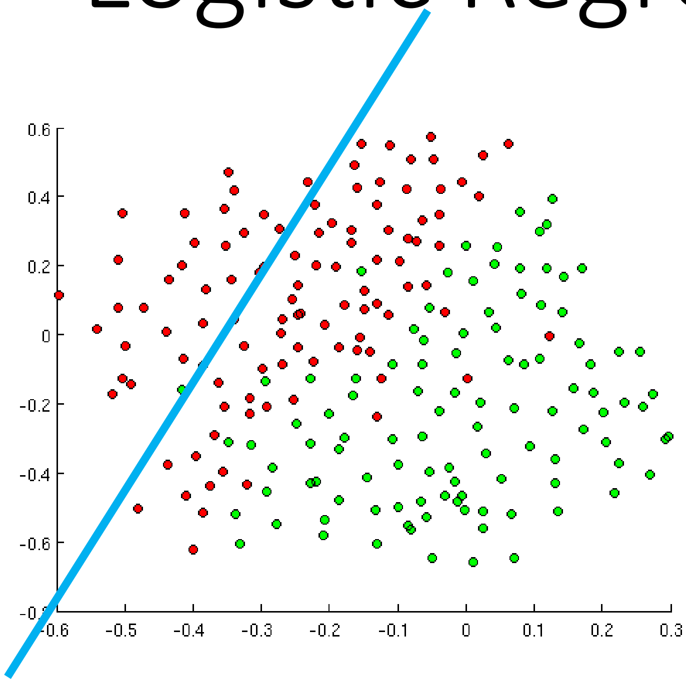


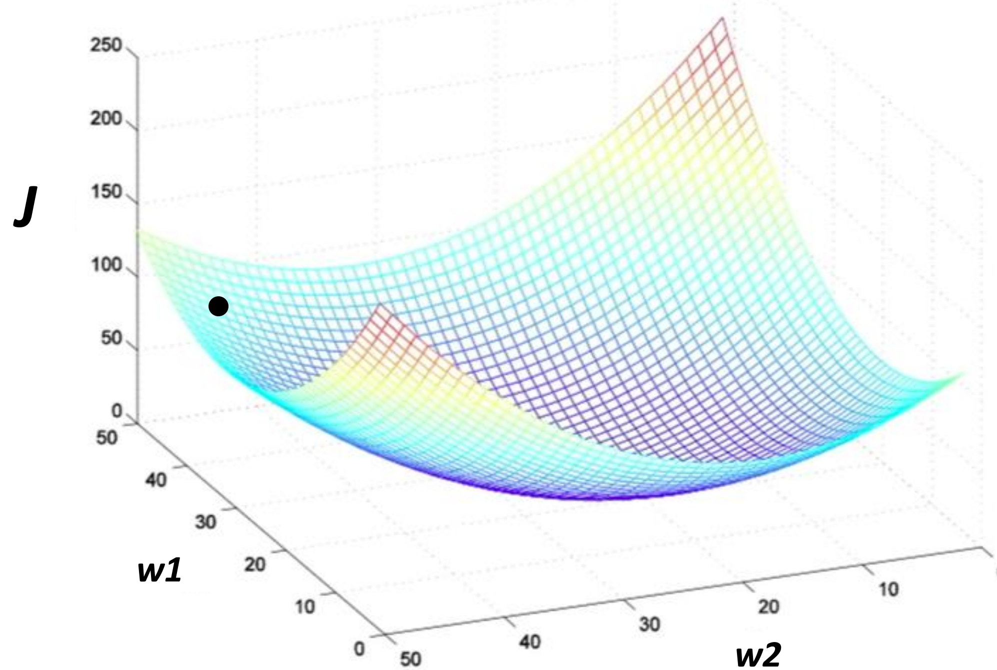


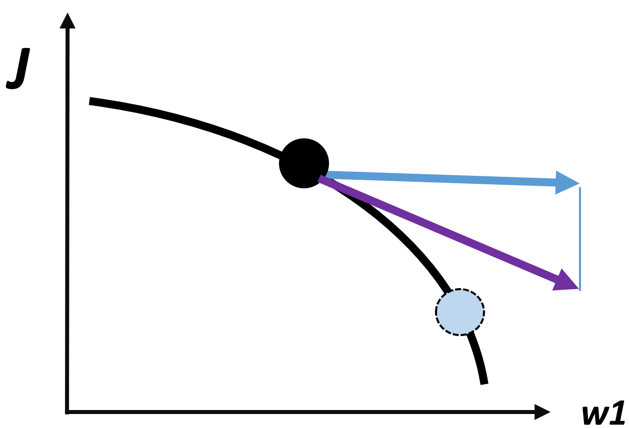
Total cost:

J(w) = Σ (y' - y)^2

Logistic Regression cont'd



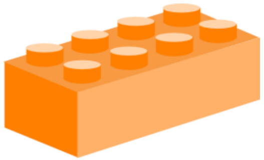


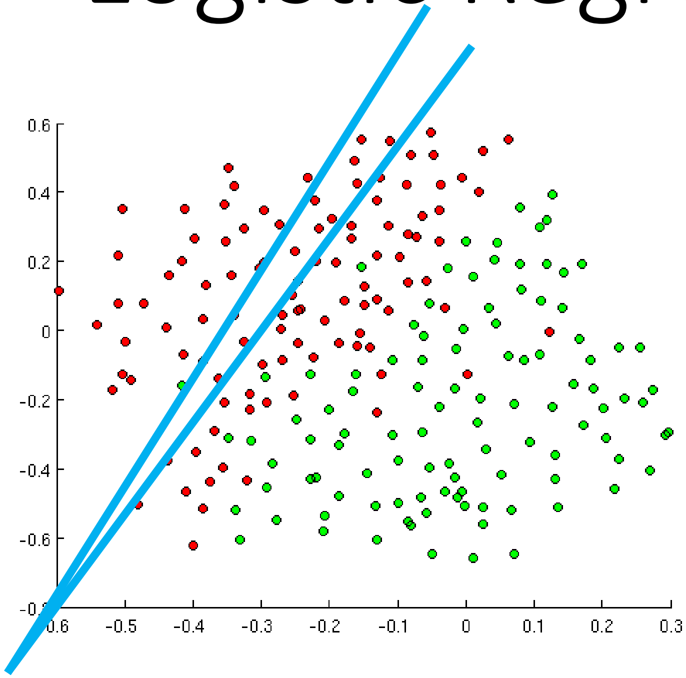


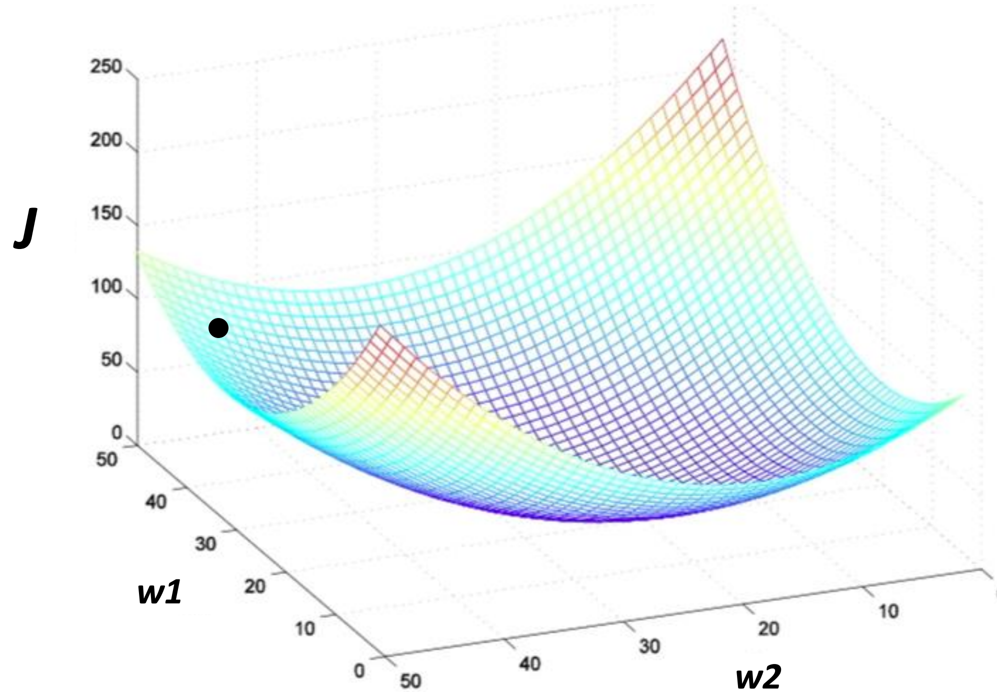
Total cost:

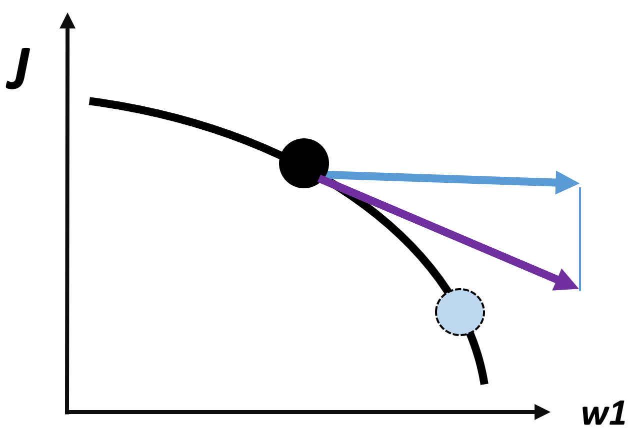
J(w) = Σ (y' - y)^2

Logistic Regression cont'd





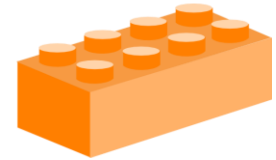


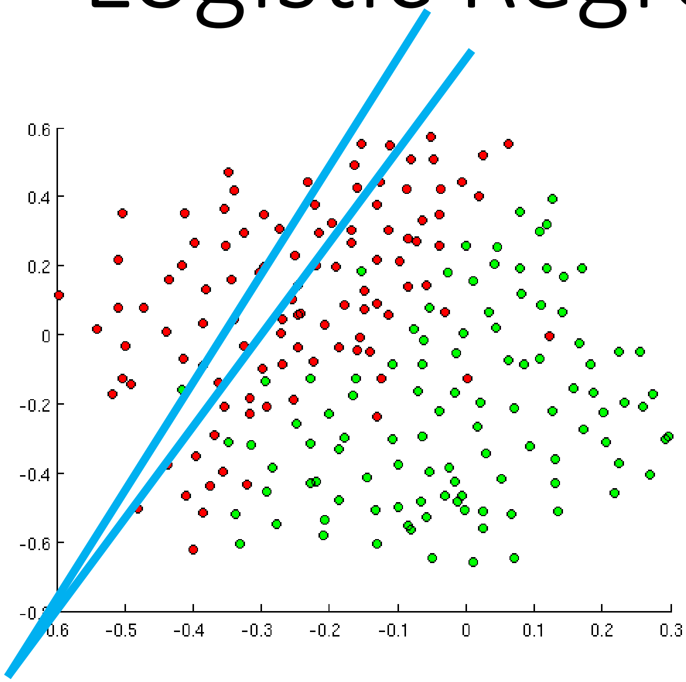


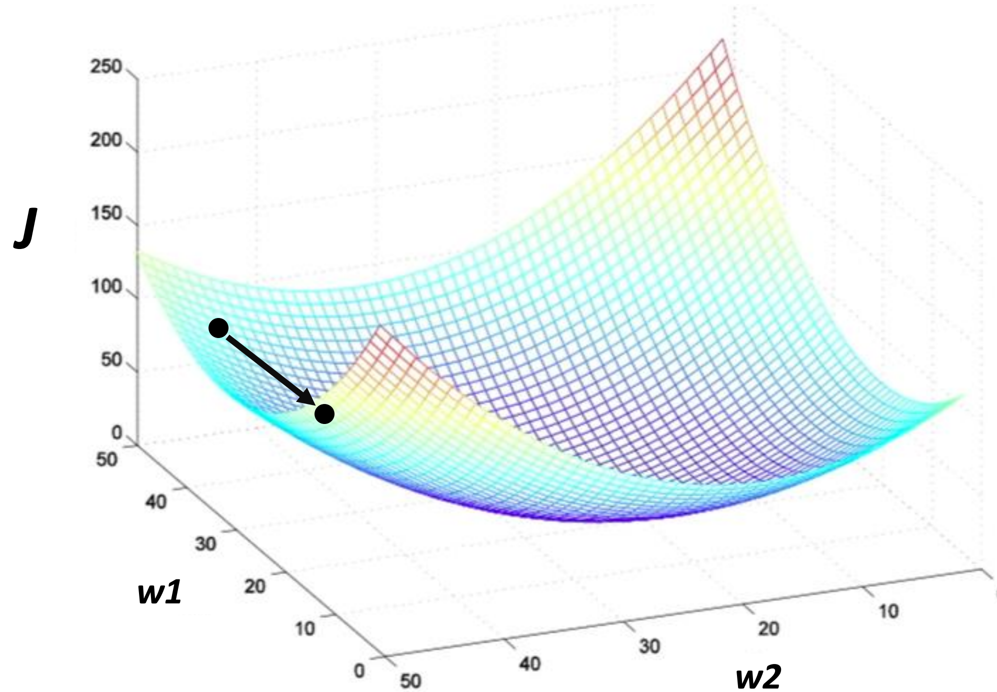
Total cost:

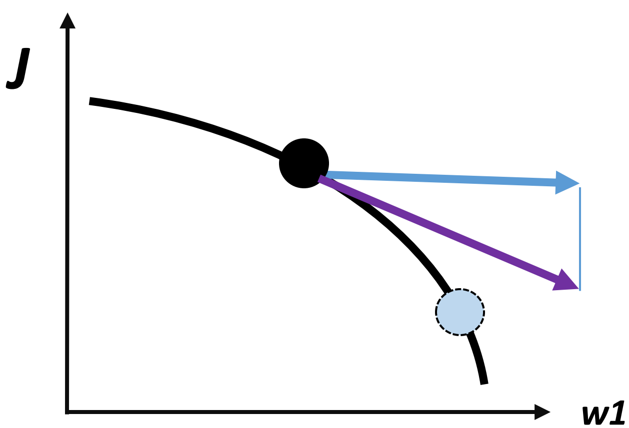
$$J(w) = \Sigma\ (y' - y)^2$$

Logistic Regression cont'd





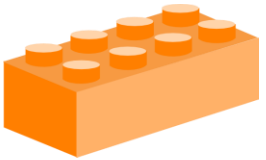


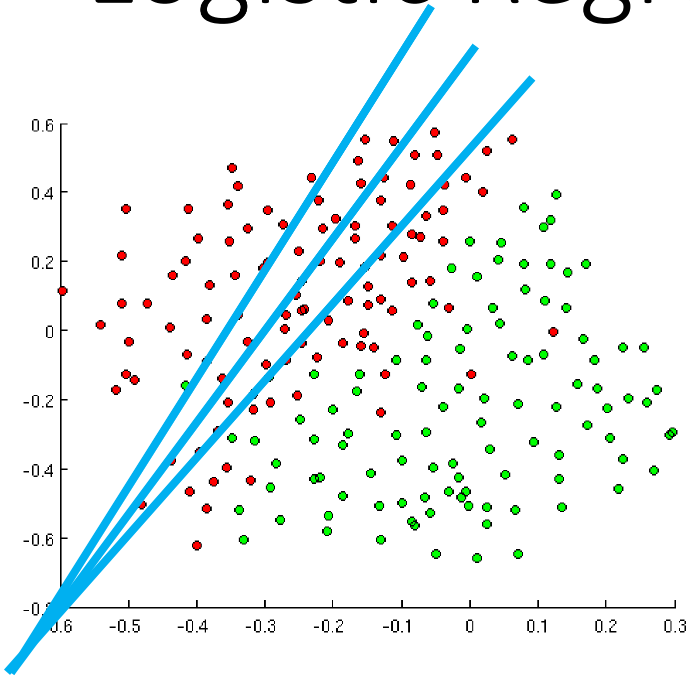


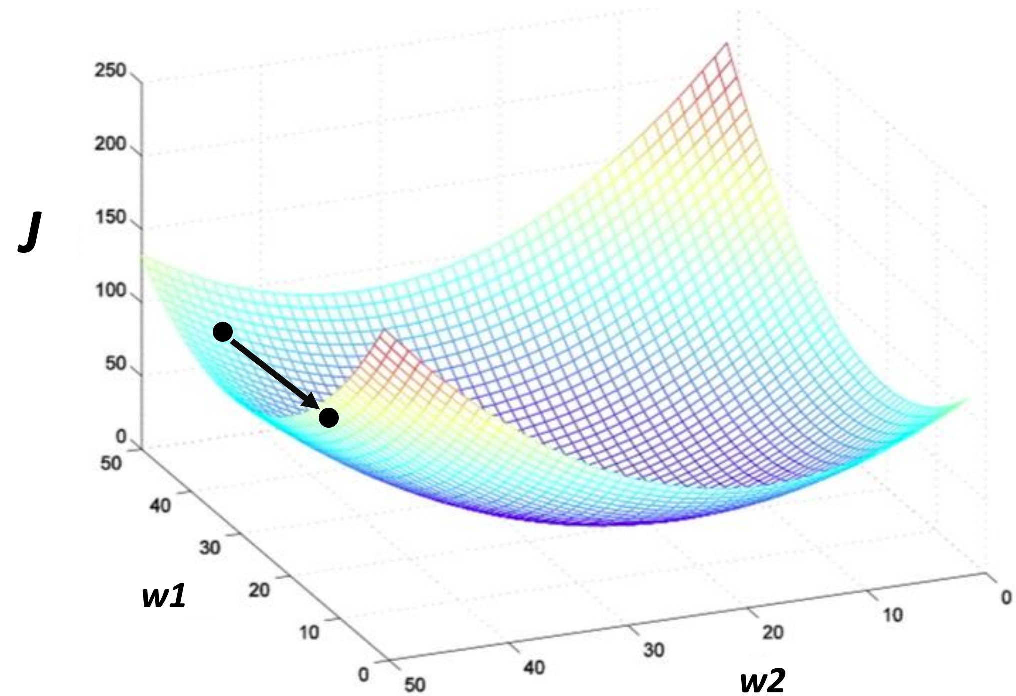
Total cost:

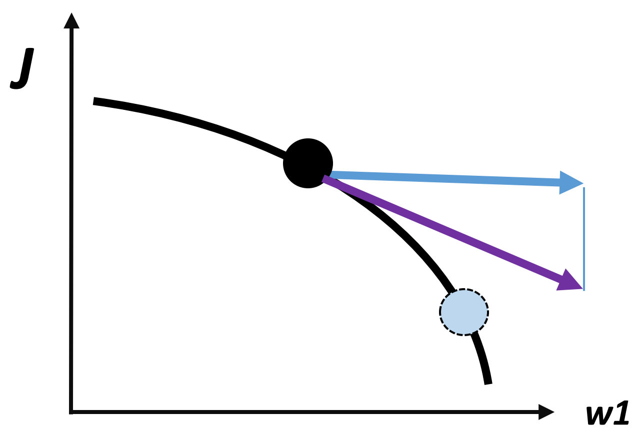
J(w) = Σ (y' - y)^2

Logistic Regression cont'd





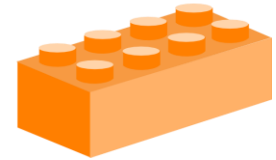


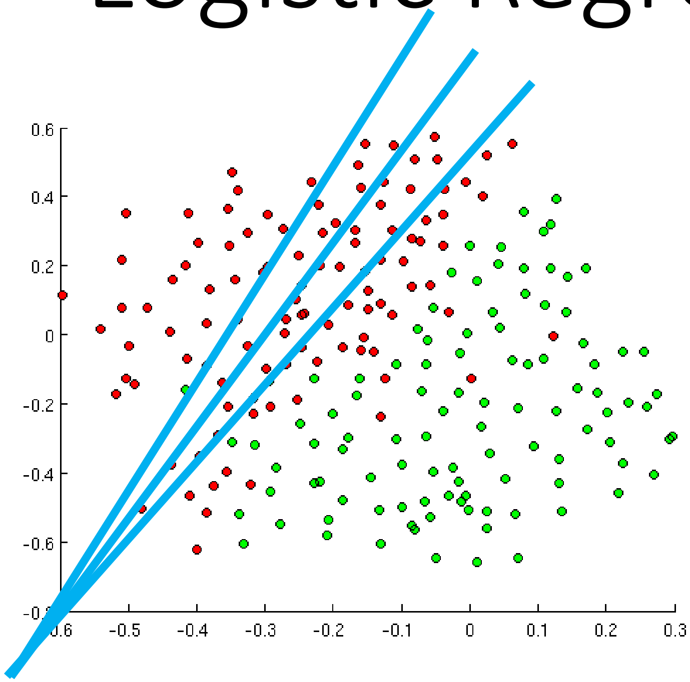


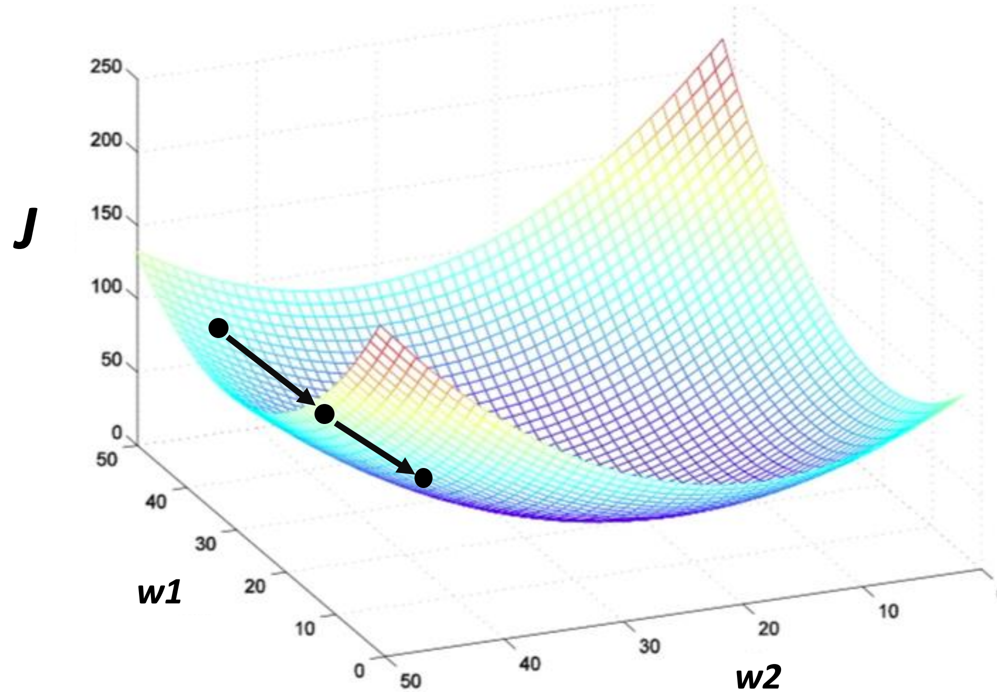
Total cost:

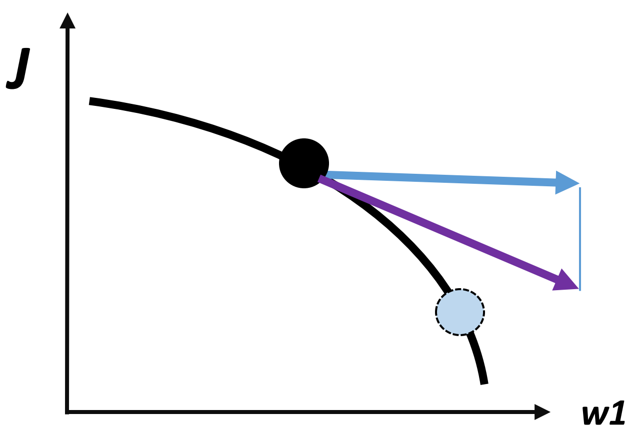
$$J(w) = \Sigma\ (y' - y)^2$$

Logistic Regression cont’d





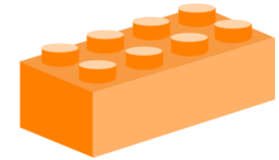


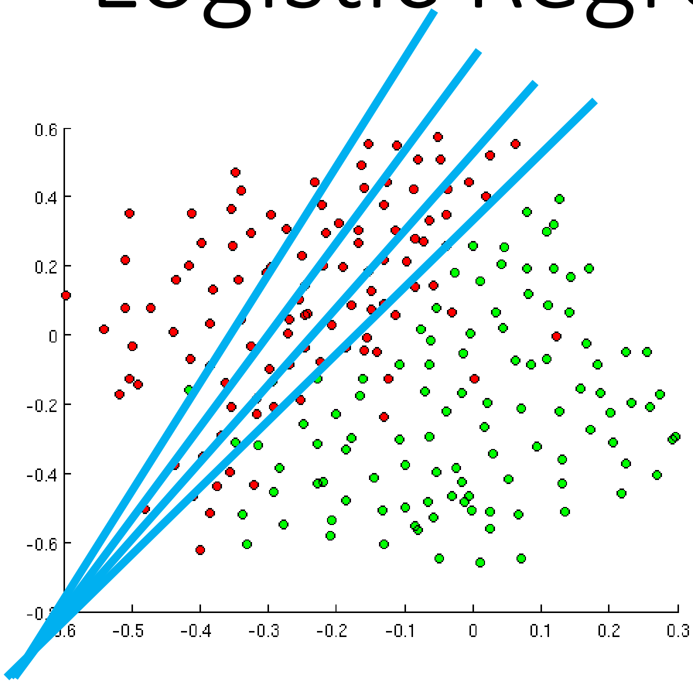


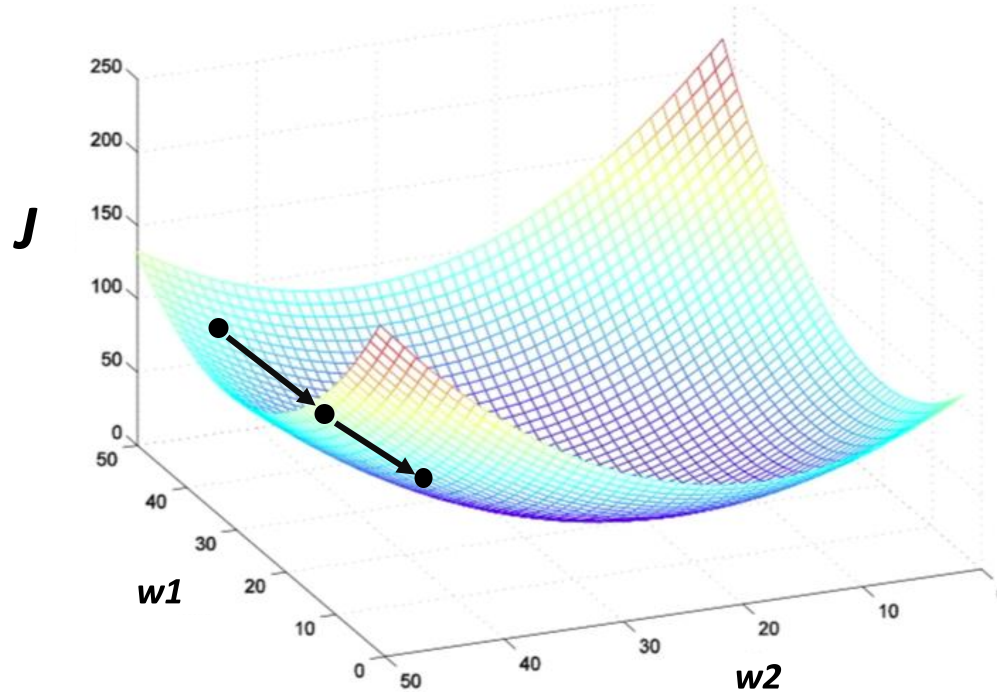
Total cost:

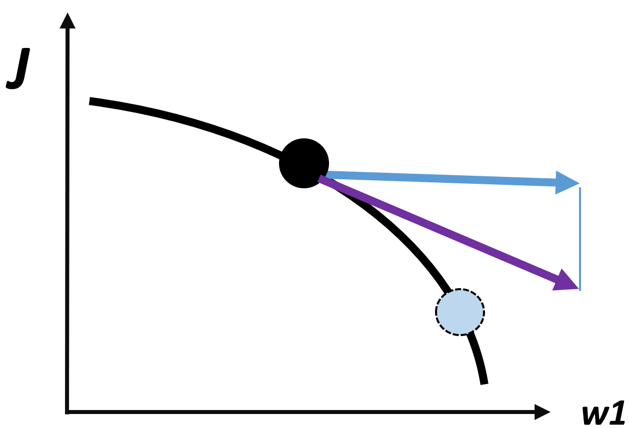
J(w) = *Σ* (y’ - y)^2

Logistic Regression cont'd





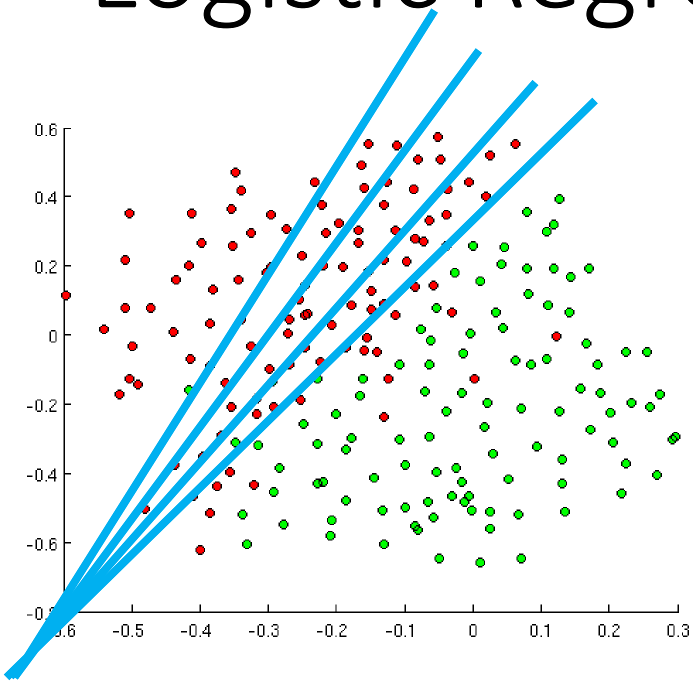


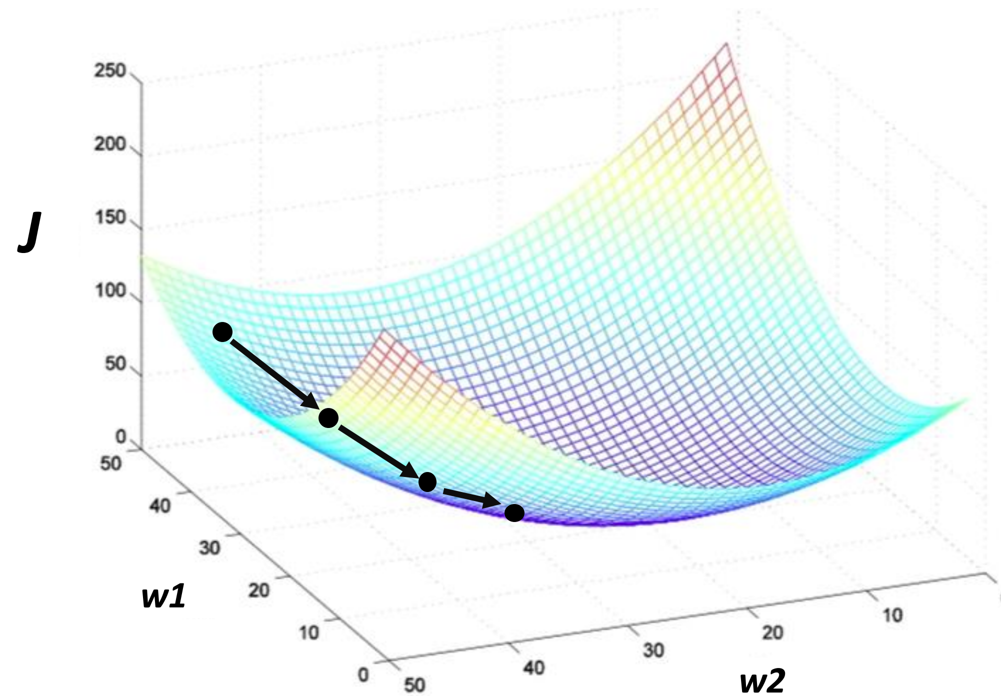


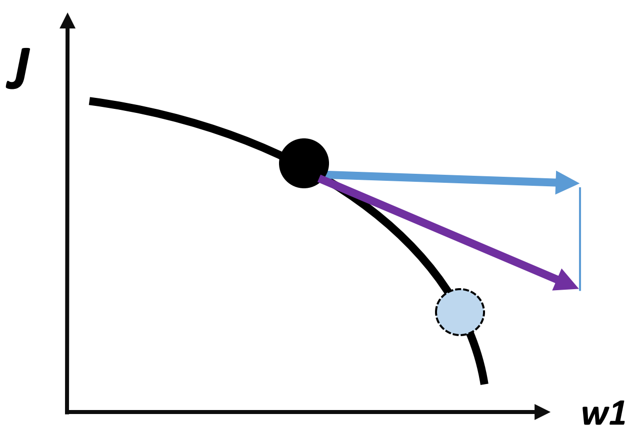
Total cost:

$$\boldsymbol{J(w)} = \Sigma\ (y' - y)^2$$

Logistic Regression cont'd



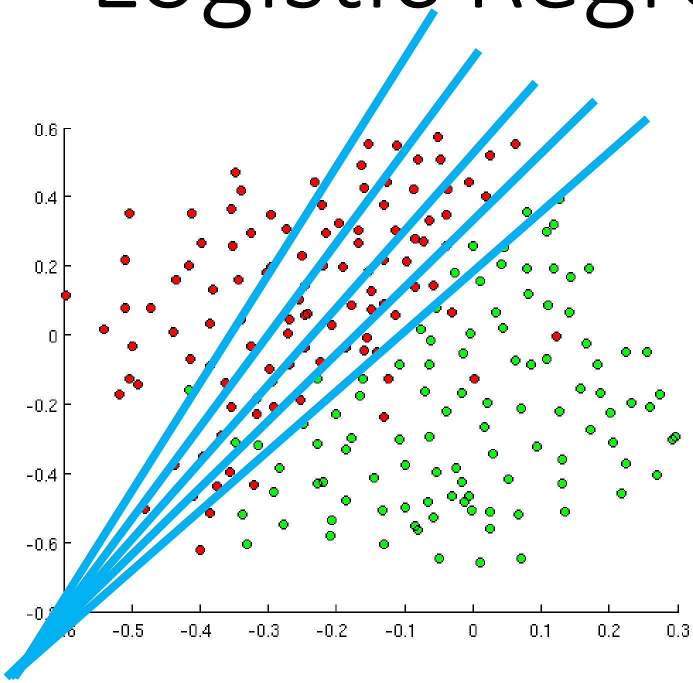


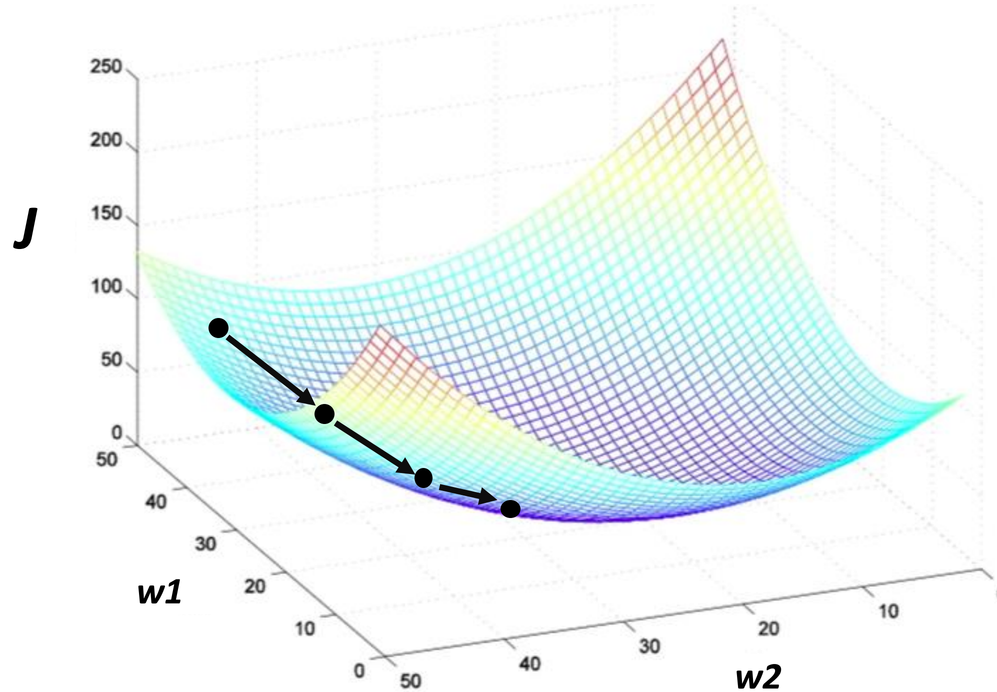


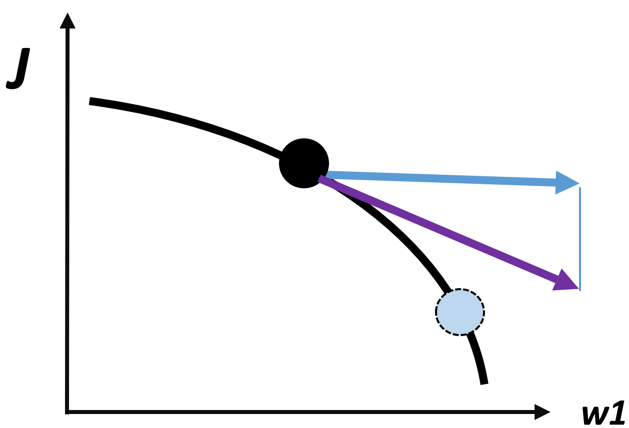
Total cost:

$$J(w) = \Sigma\ (y' - y)^2$$

Logistic Regression cont’d



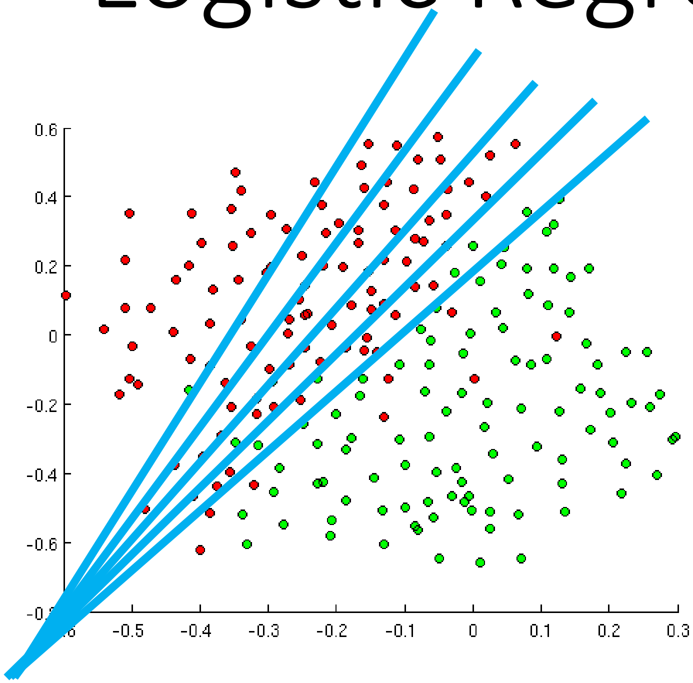


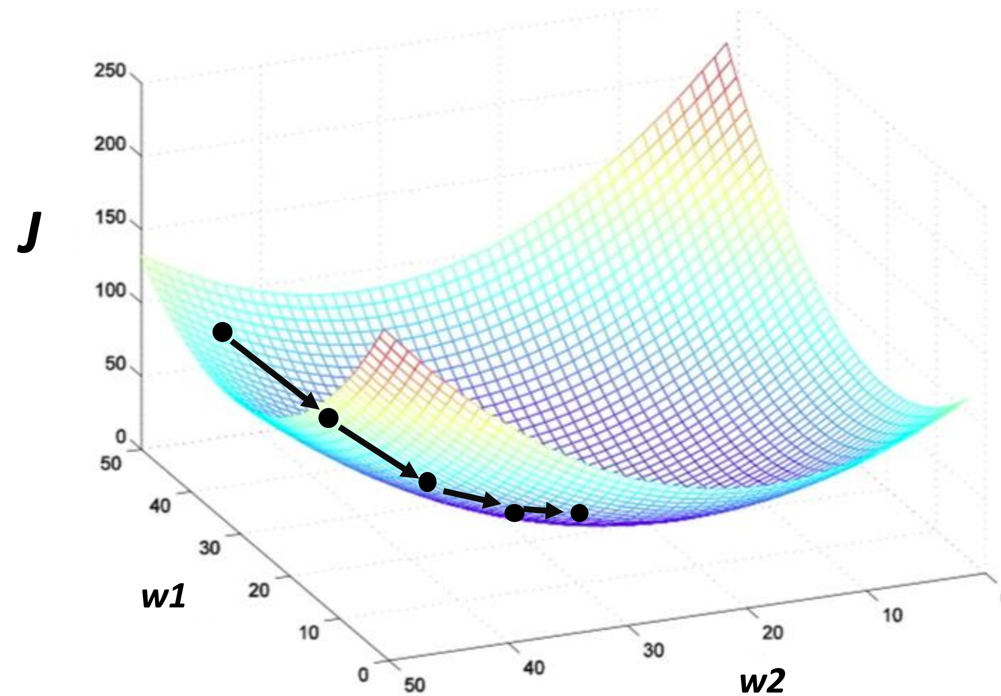


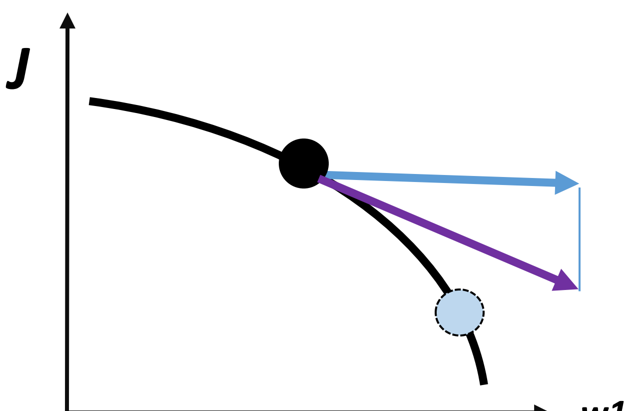
Total cost:

$$J(w) = \Sigma\ (y' - y)^2$$

Logistic Regression cont'd



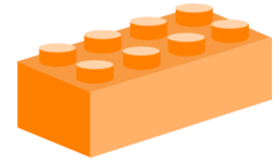


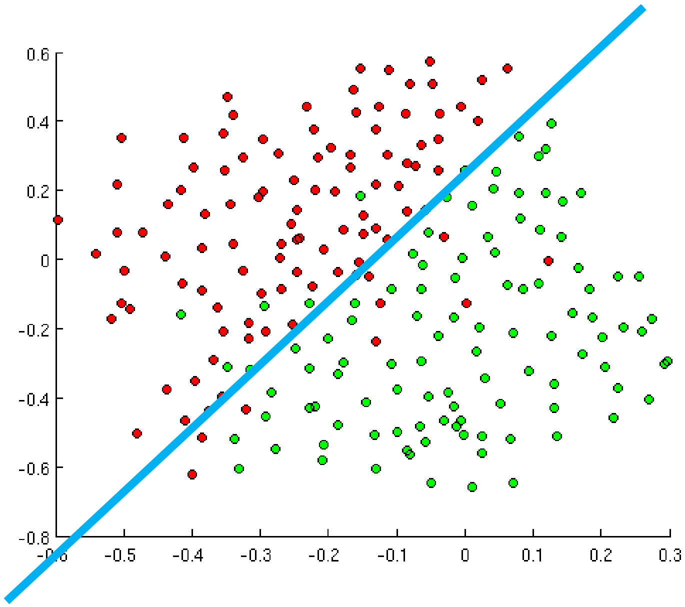


Total cost:

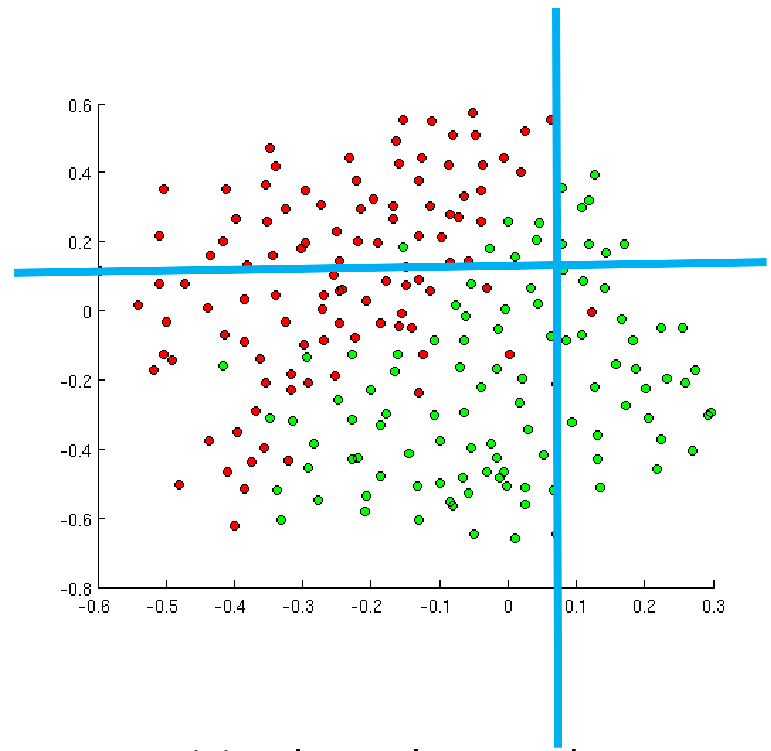
$$J(w) = \Sigma\,(y' - y)^2$$

Logistic Regression cont’d





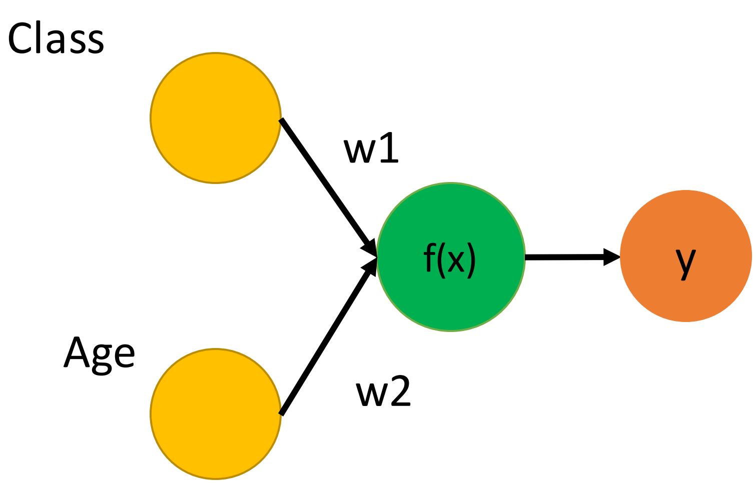
Decision boundary made by a linear model (logistic regression)



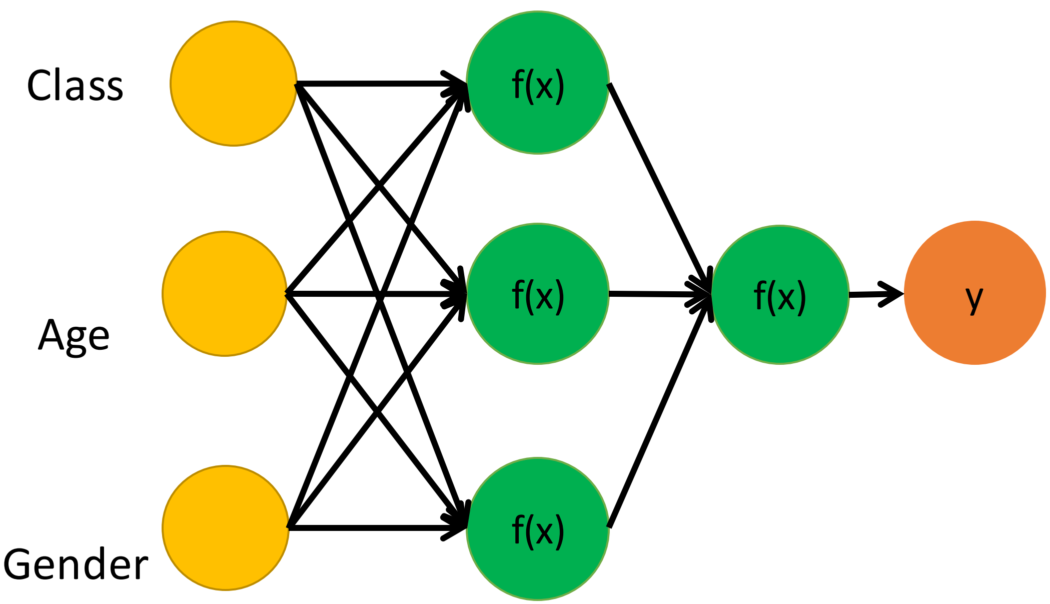
Decision boundary made by a decision tree.

Neural Network

- Adds multiple layer of interconnected regression node.
- Following the same way of training.



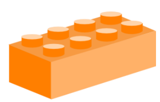
A logistic regression model.

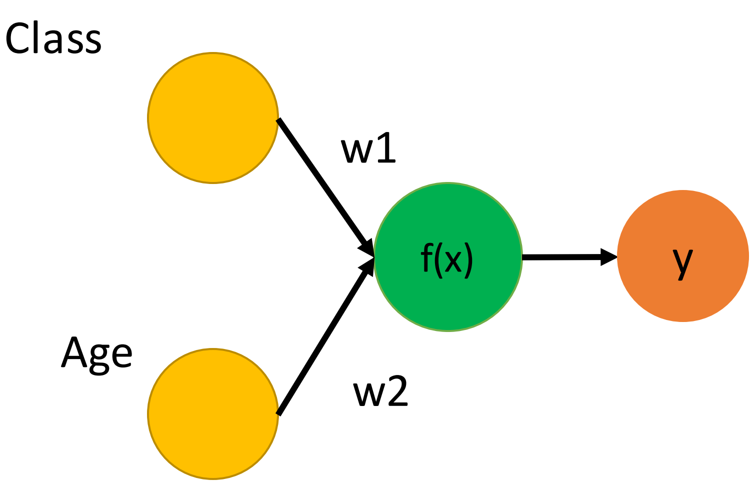


A neural network model.

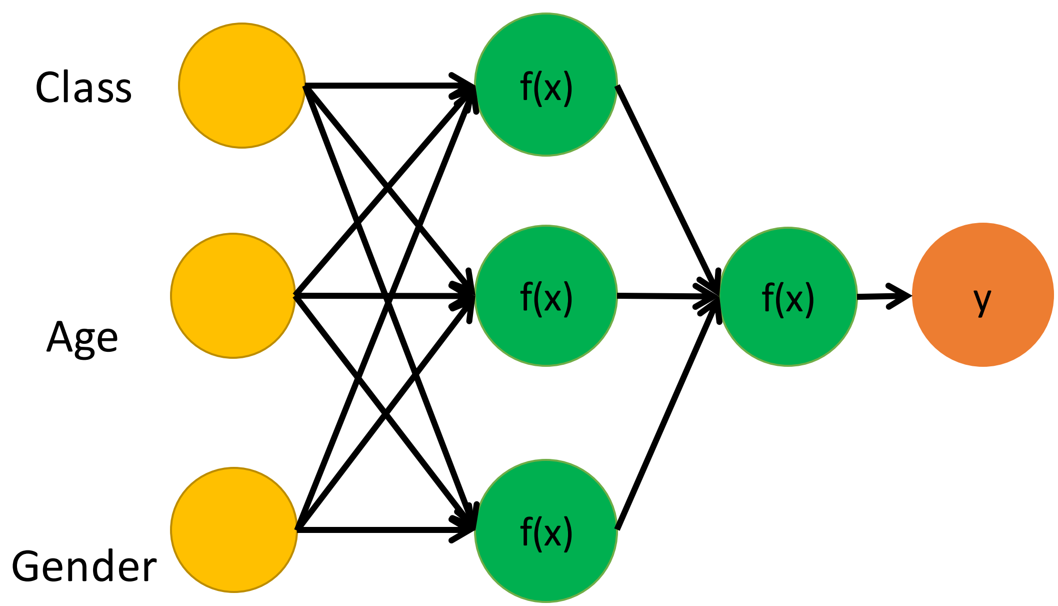
Neural Network

- Adds multiple layer of interconnected regression node.
- Following the same way of training.





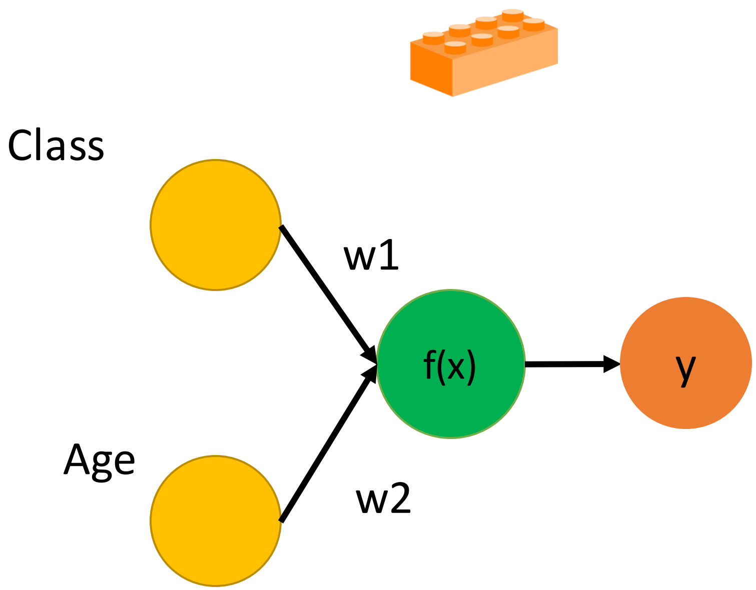
A logistic regression model.



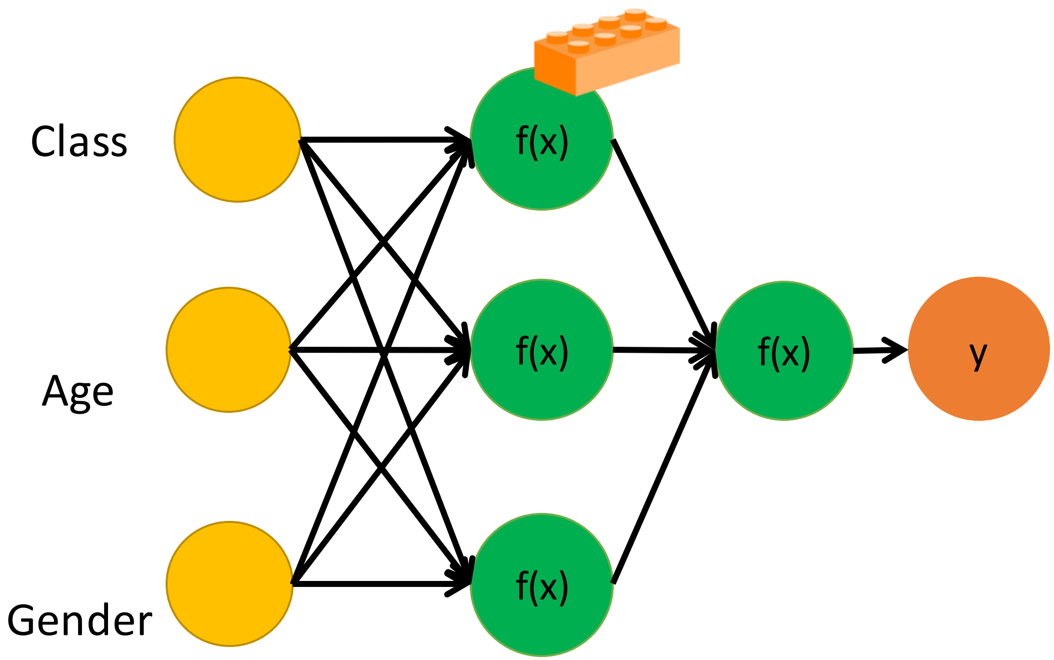
A neural network model.

Neural Network

- Adds multiple layer of interconnected regression node.
- Following the same way of training.



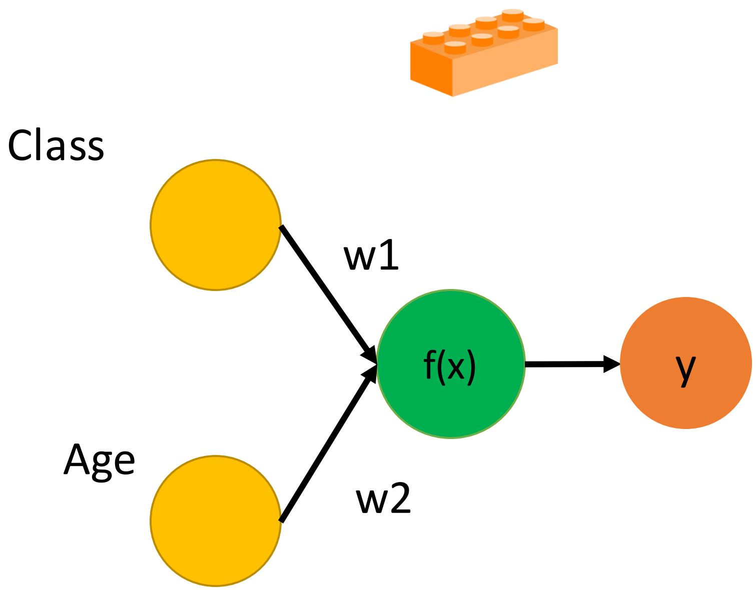
A logistic regression model.



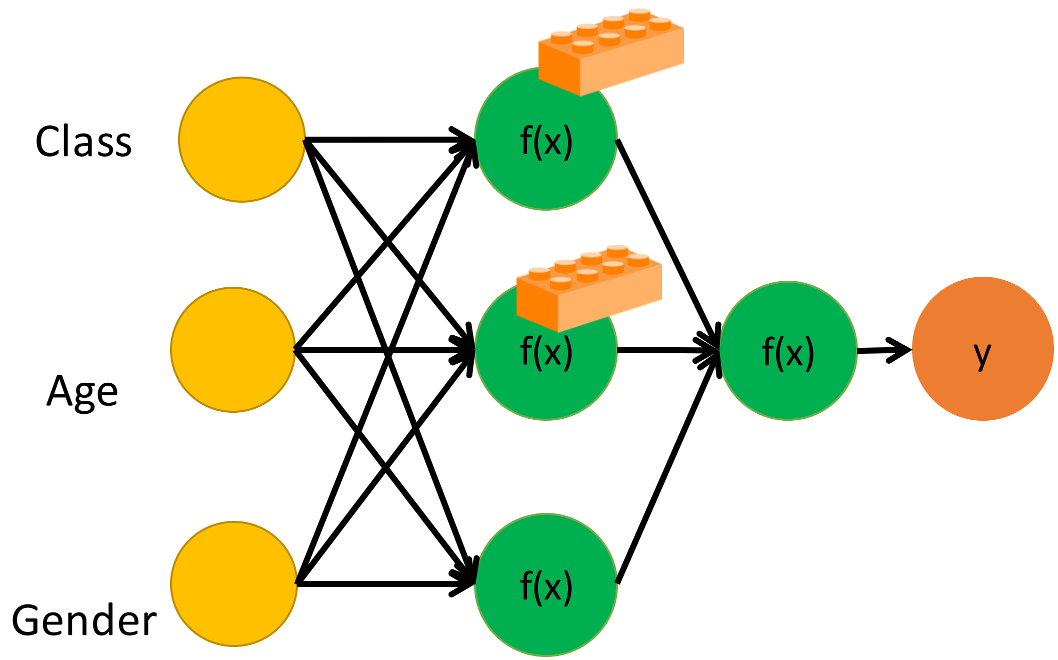
A neural network model.

Neural Network

- Adds multiple layer of interconnected regression node.
- Following the same way of training.



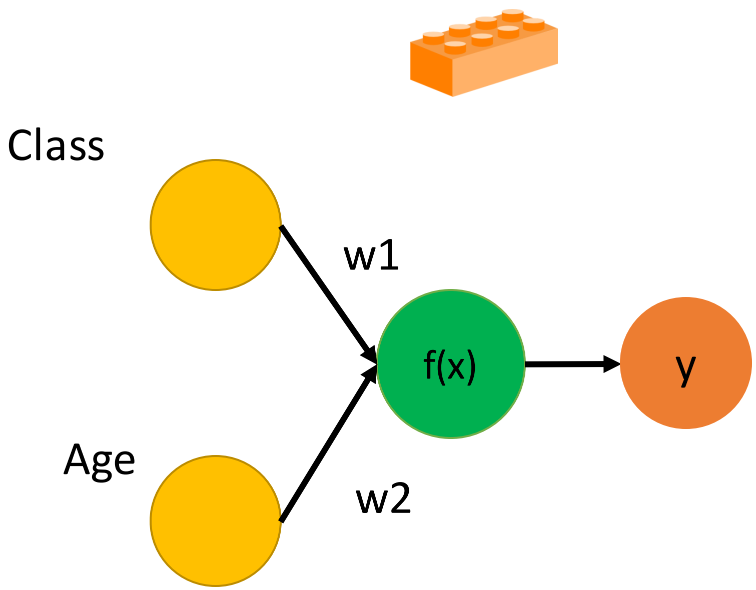
A logistic regression model.



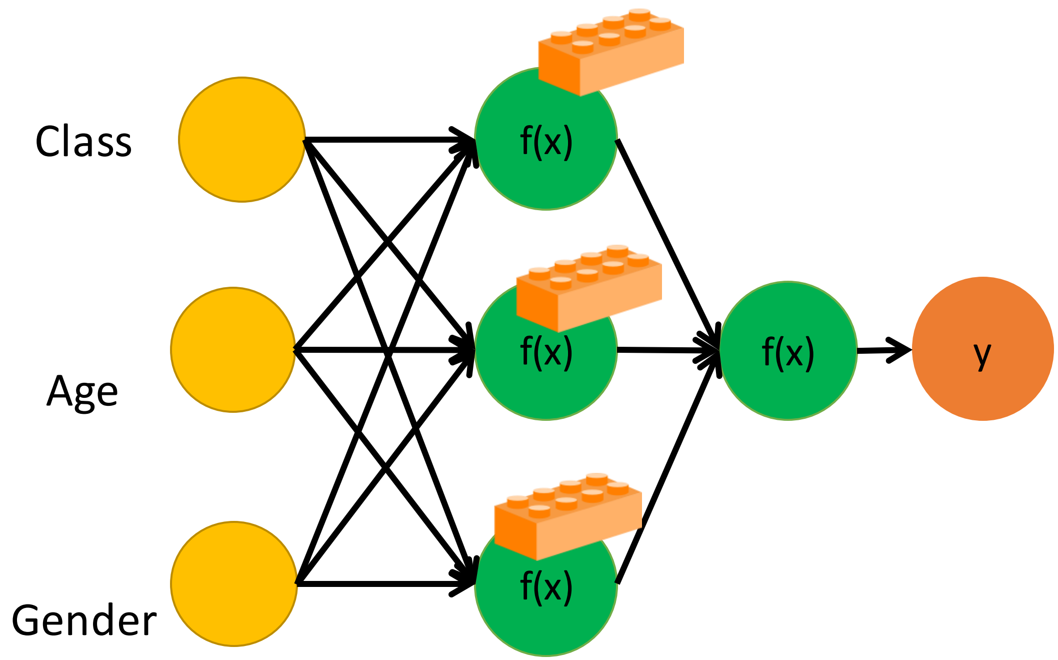
A neural network model.

Neural Network

- Adds multiple layer of interconnected regression node.
- Following the same way of training.



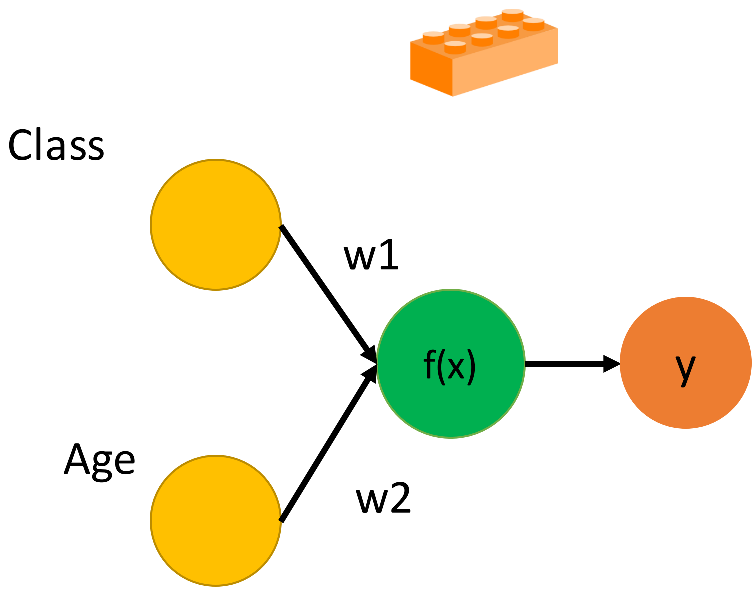
A logistic regression model.



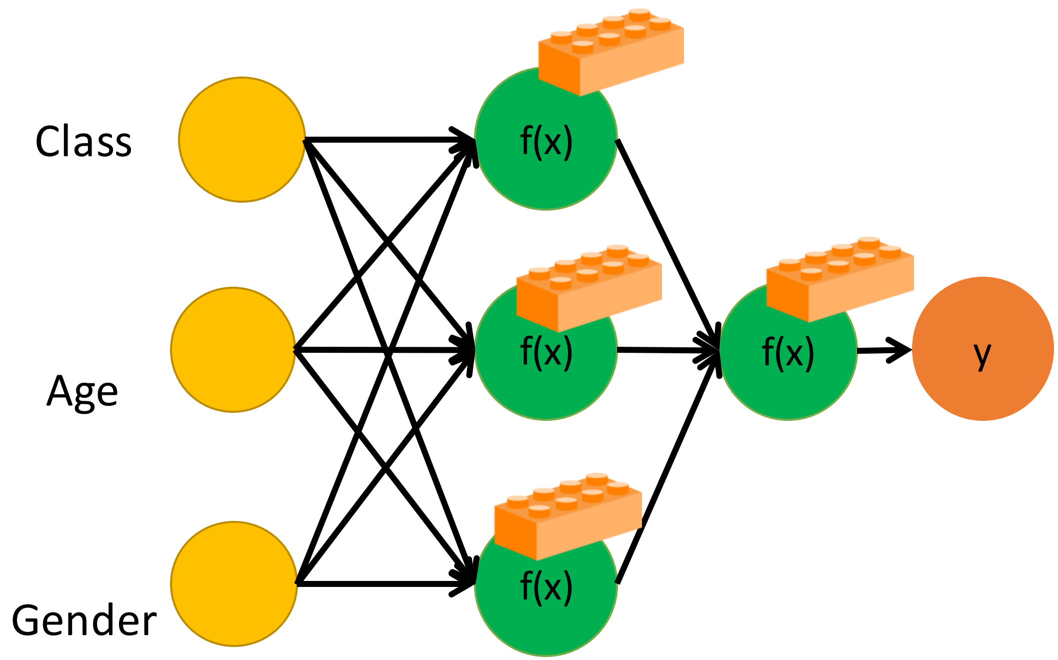
A neural network model.

Neural Network

- Adds multiple layer of interconnected regression node.
- Following the same way of training.



A logistic regression model.

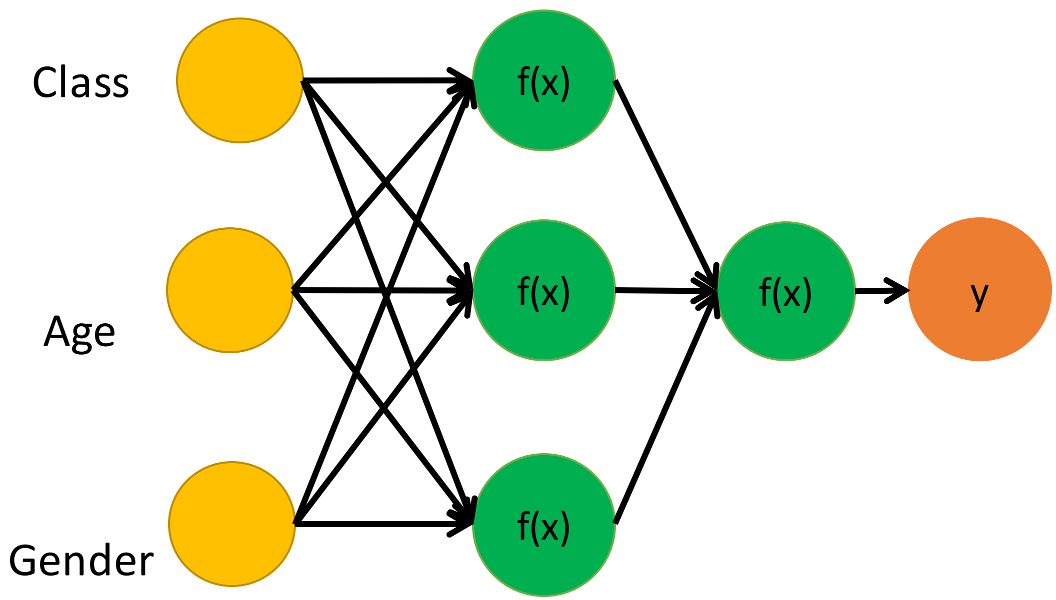


A neural network model.

Neural Network

Passenger Class	Gender	age	Survived
1	1	29	1
1	2	2	0
2	2	21	1
2	1	19	1

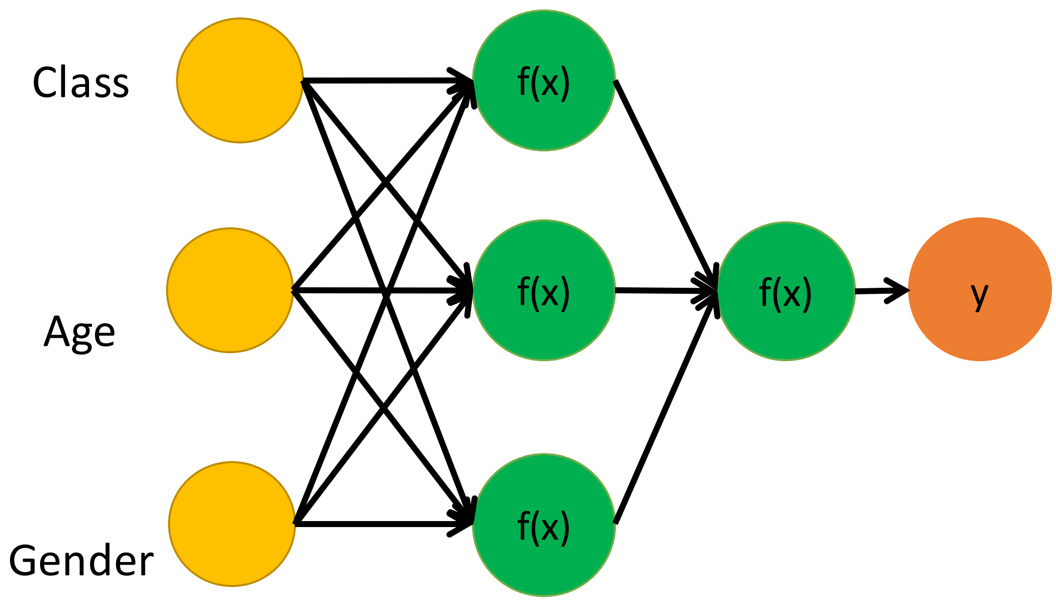
Randomly initialize weights.



Neural Network

Passenger Class	Gender	age	Survived
1	1	29	1
1	2	2	0
2	2	21	1
2	1	19	1

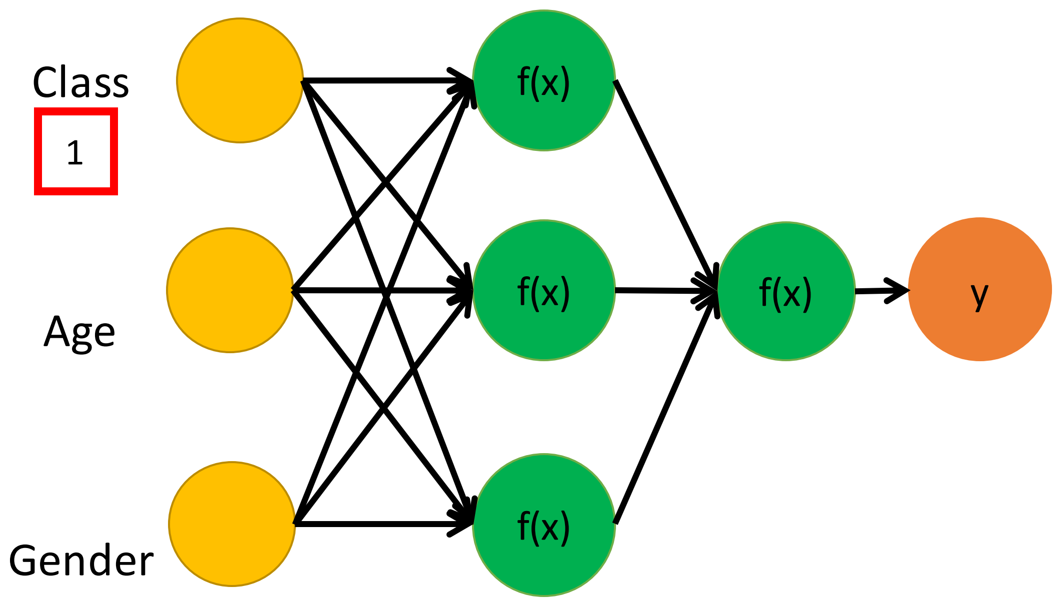
Randomly initialize weights.



Neural Network

Passenger Class	Gender	age	Survived
1	1	29	1
1	2	2	0
2	2	21	1
2	1	19	1

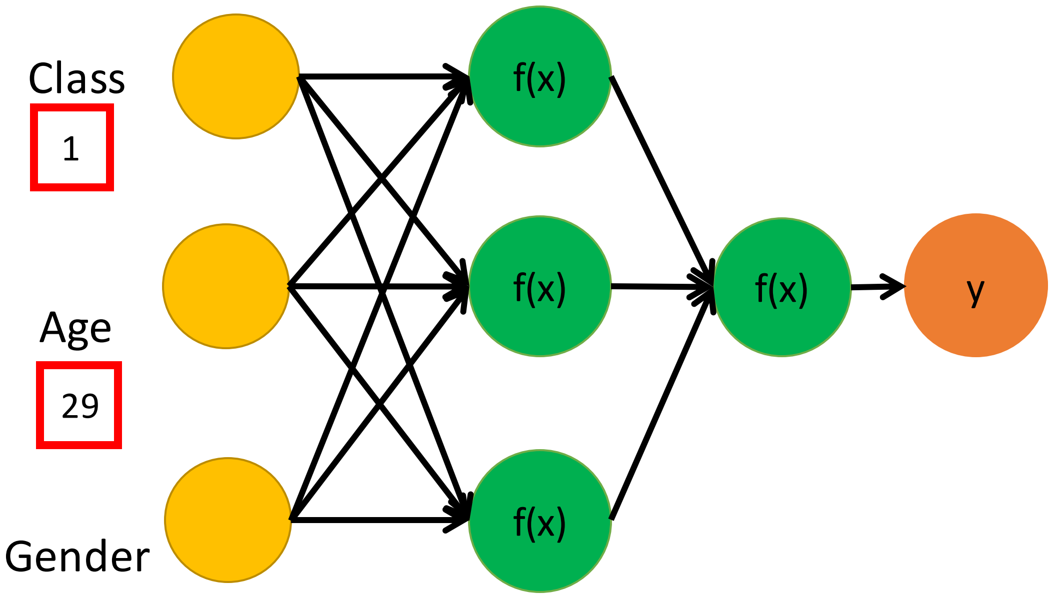
Randomly initialize weights.



Neural Network

Passenger Class	Gender	age	Survived
1	1	29	1
1	2	2	0
2	2	21	1
2	1	19	1

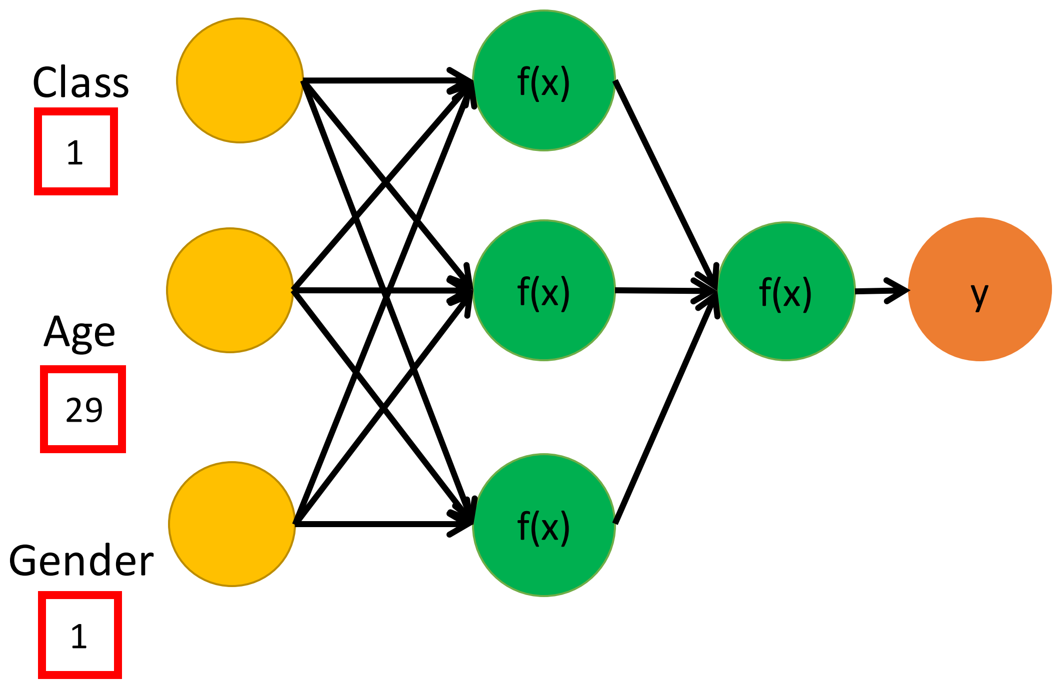
Randomly initialize weights.



Neural Network

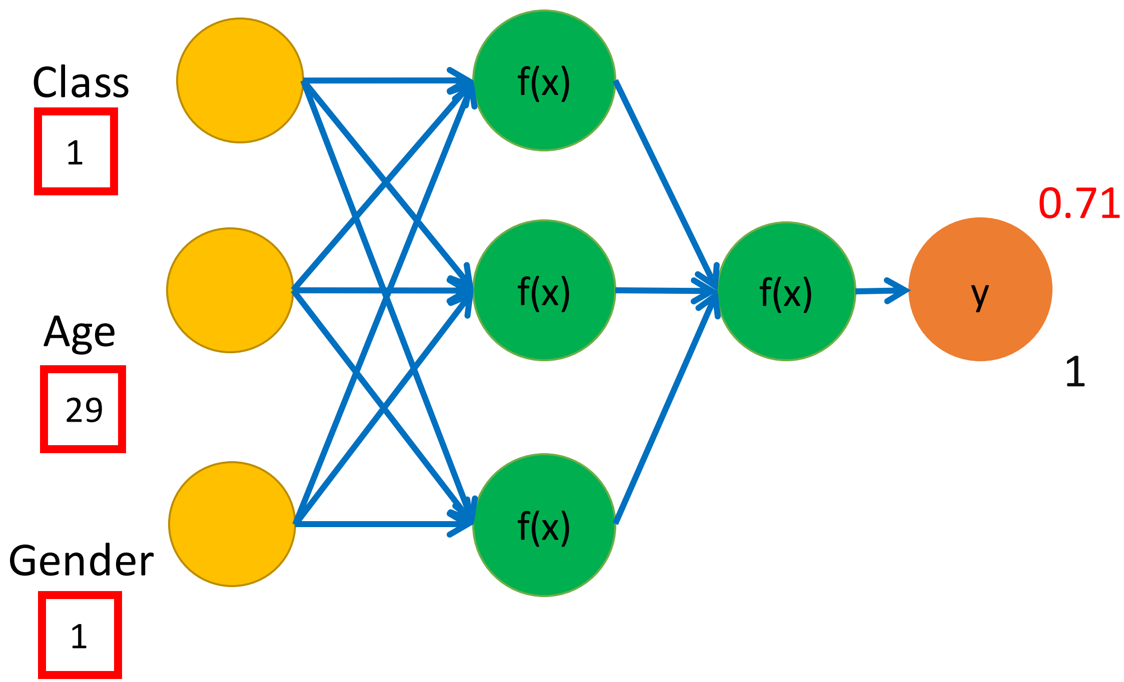
Passenger Class	Gender	age	Survived
1	1	29	1
1	2	2	0
2	2	21	1
2	1	19	1

Randomly initialize weights.



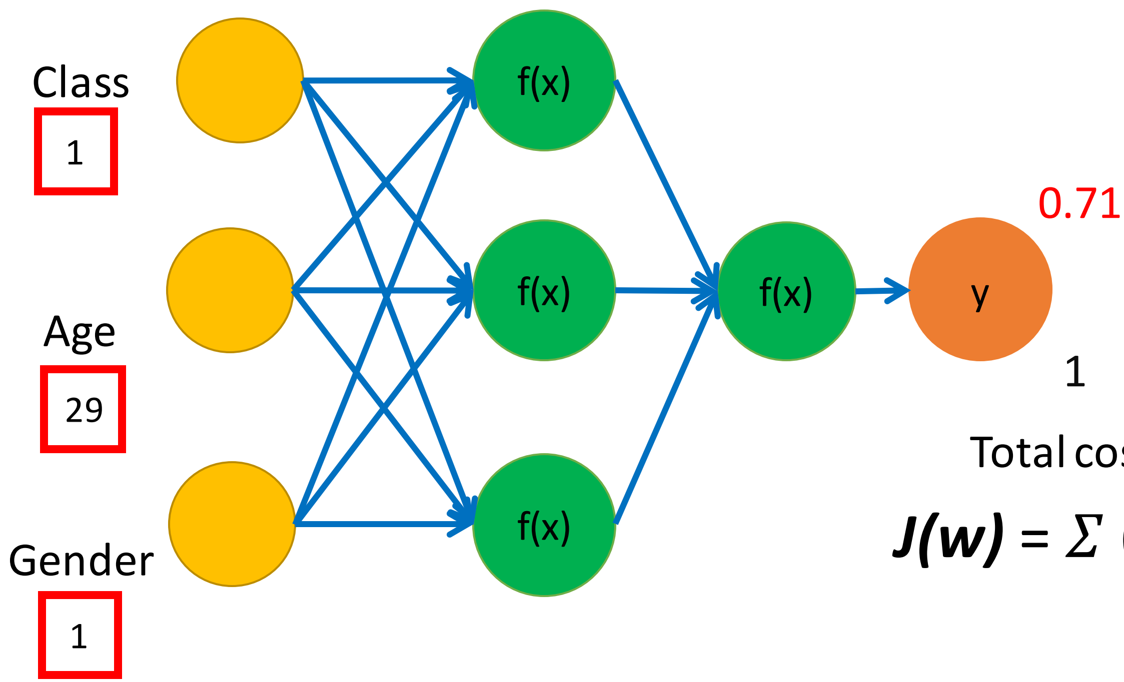
Neural Network

Passenger Class	Gender	age	Survived
1	1	29	1
1	2	2	0
2	2	21	1
2	1	19	1



Neural Network

Passenger Class	Gender	age	Survived
1	1	29	1
1	2	2	0
2	2	21	1
2	1	19	1

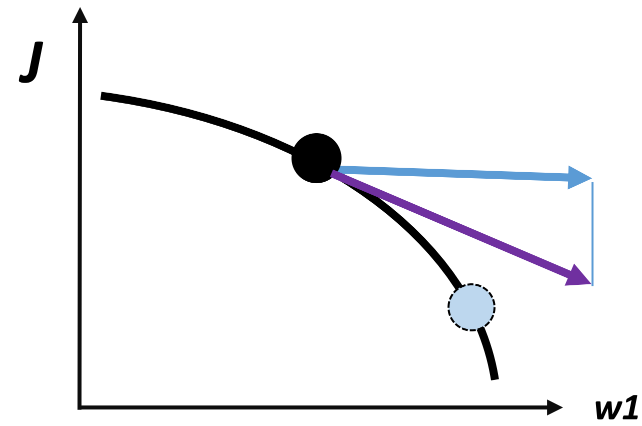


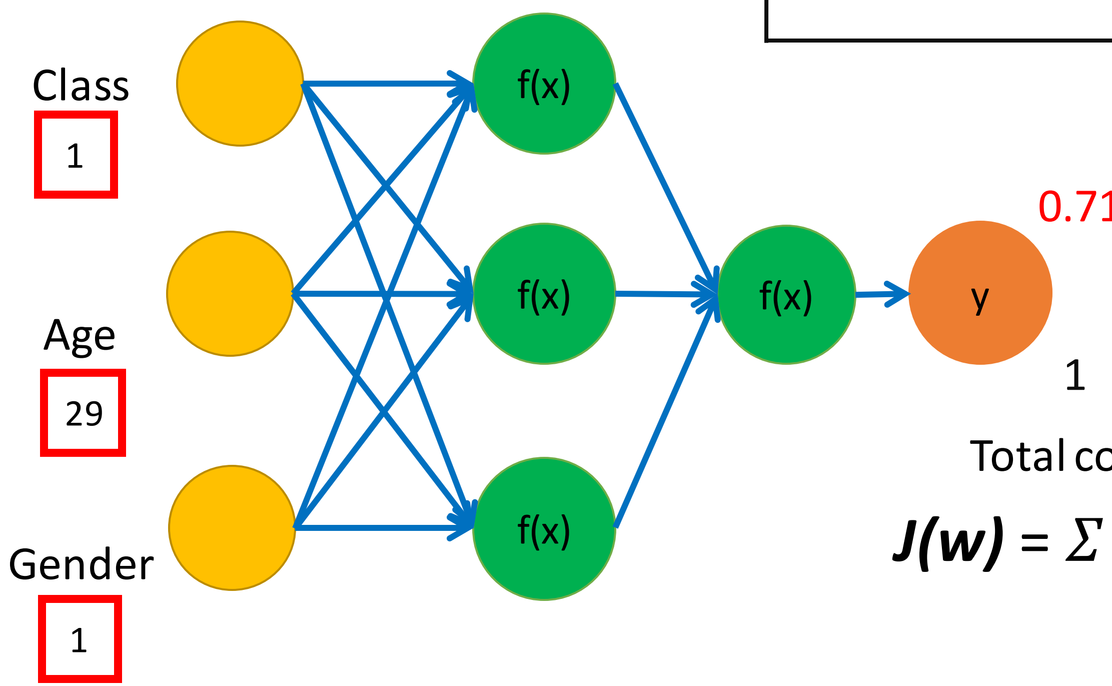
Total cost:

$$J(w) = \Sigma\,(y' - y)^2$$

Neural Network

Passenger Class	Gender	age	Survived
1	1	29	1
1	2	2	0
2	2	21	1
2	1	19	1



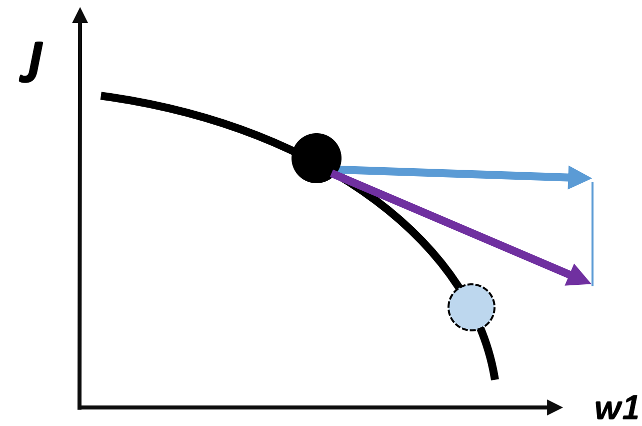


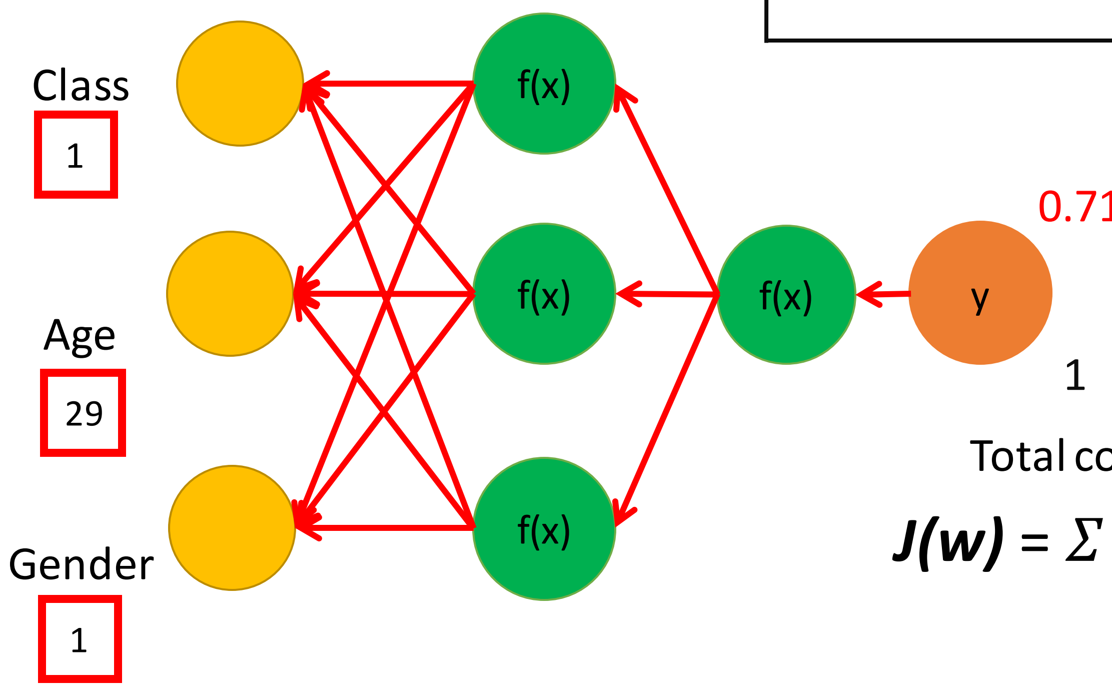
Total cost:

$$J(w) = \Sigma\, (y' - y)^2$$

Neural Network

Passenger Class	Gender	age	Survived
1	1	29	1
1	2	2	0
2	2	21	1
2	1	19	1



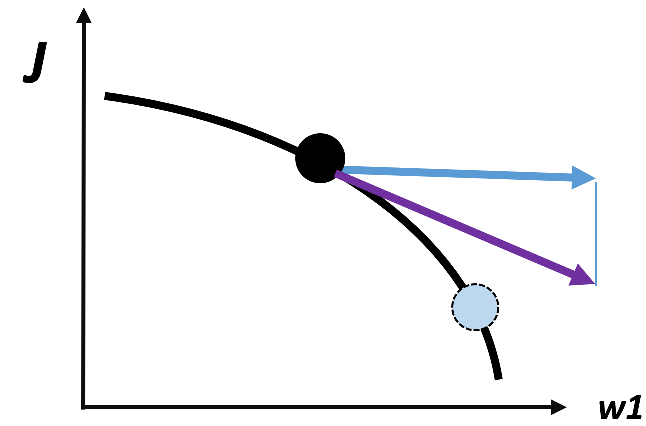


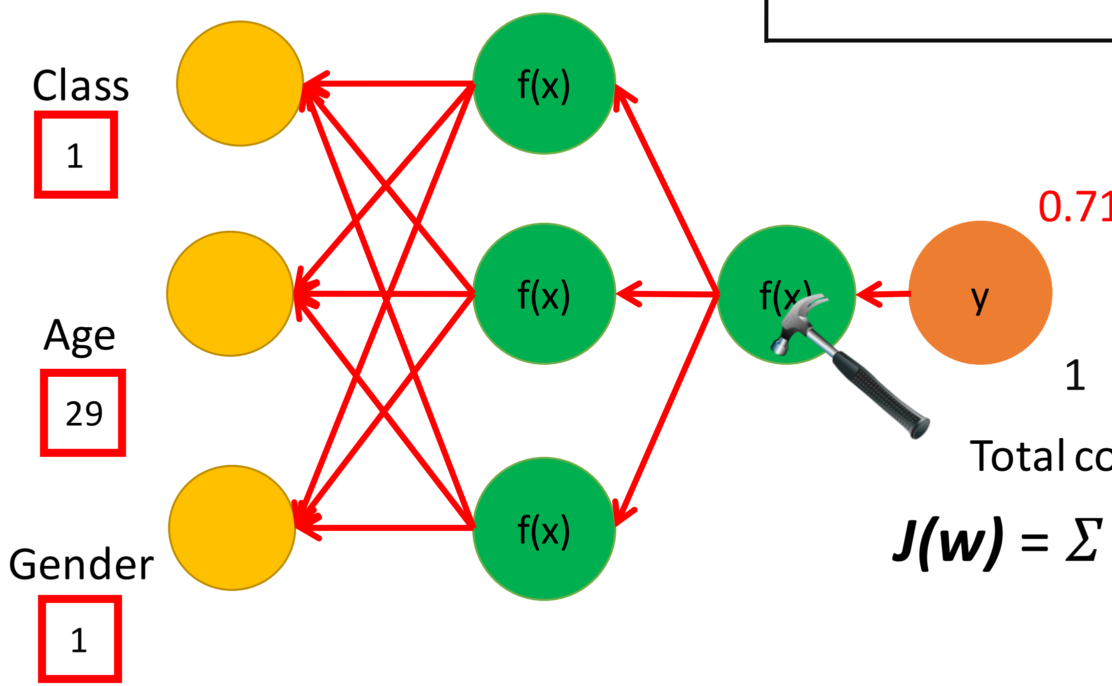
Total cost:

$$J(w) = \Sigma\ (y' - y)^2$$

Neural Network

Passenger Class	Gender	age	Survived
1	1	29	1
1	2	2	0
2	2	21	1
2	1	19	1



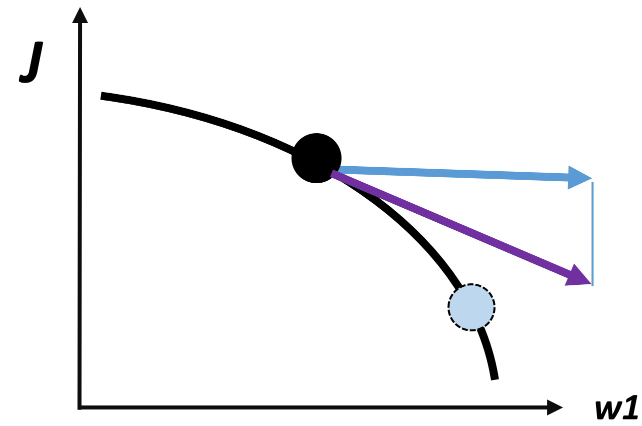


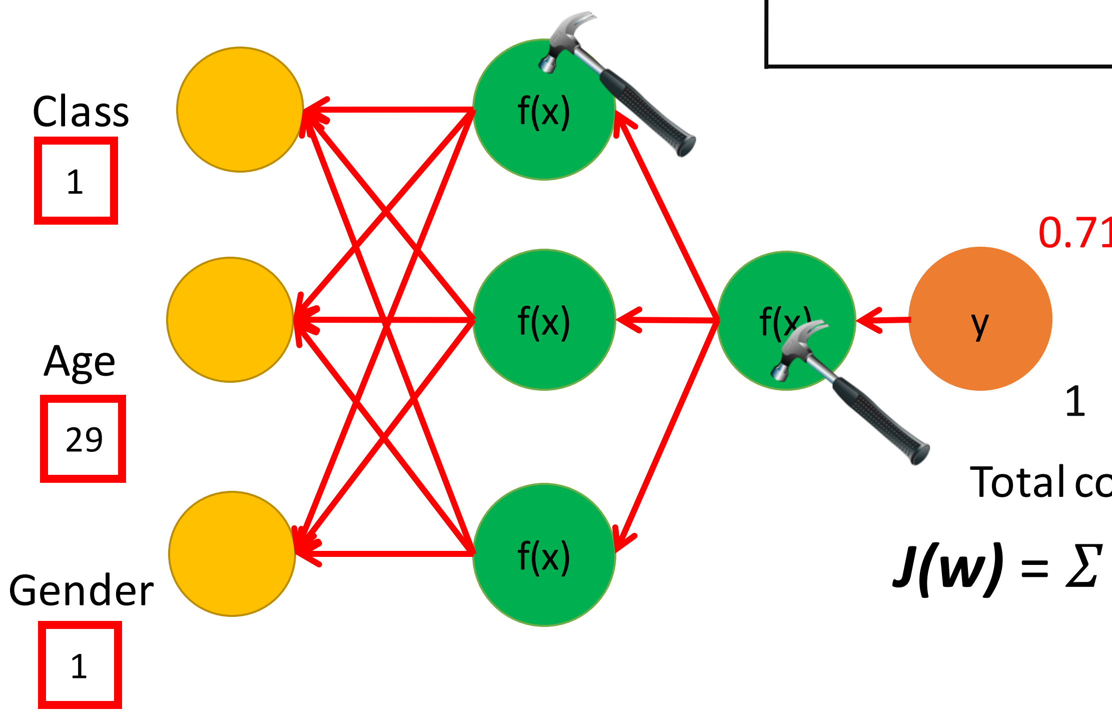
Total cost:

$$J(w) = \Sigma \, (y' - y)^2$$

Neural Network

Passenger Class	Gender	age	Survived
1	1	29	1
1	2	2	0
2	2	21	1
2	1	19	1



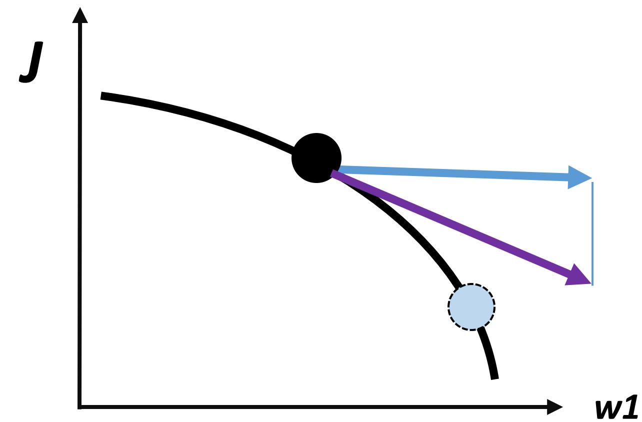


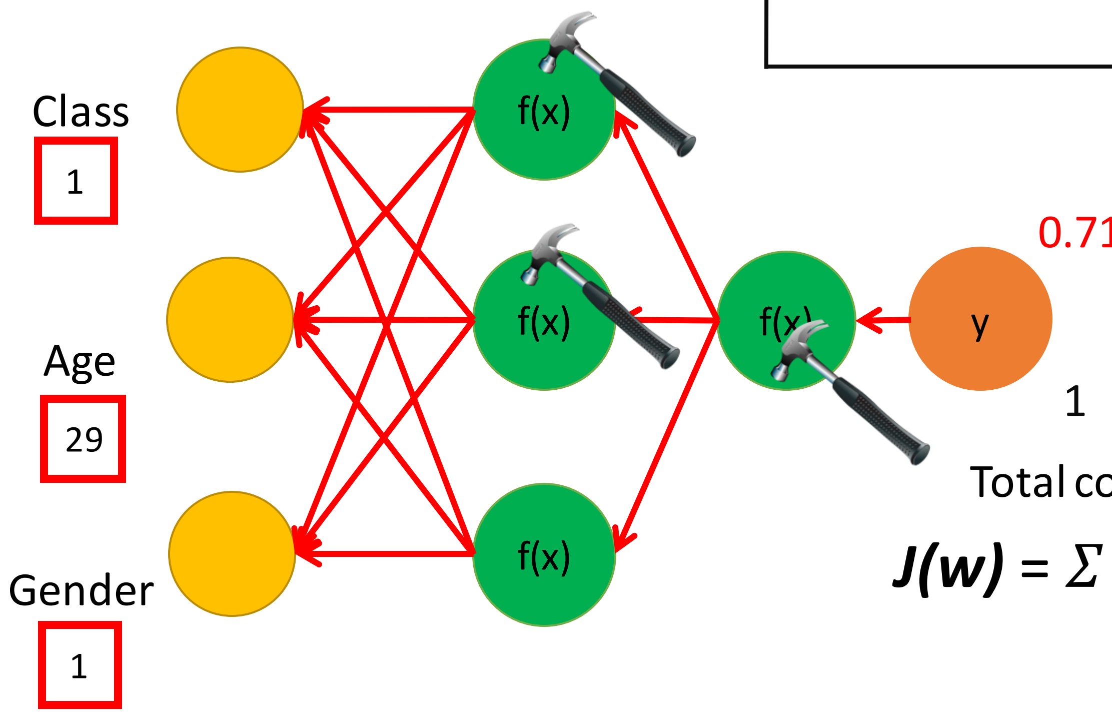
Total cost:

$$J(w) = \Sigma (y' - y)^2$$

Neural Network

Passenger Class	Gender	age	Survived
1	1	29	1
1	2	2	0
2	2	21	1
2	1	19	1



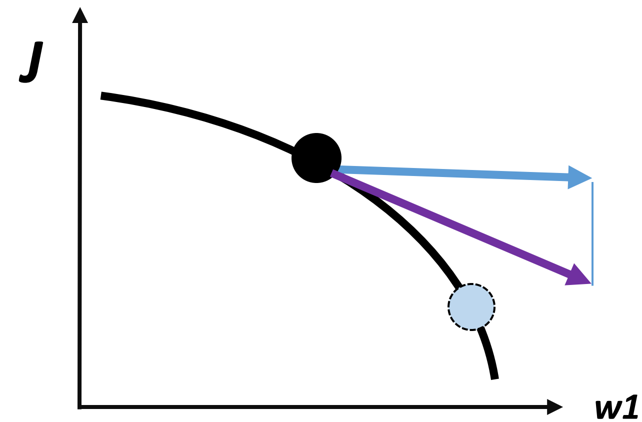


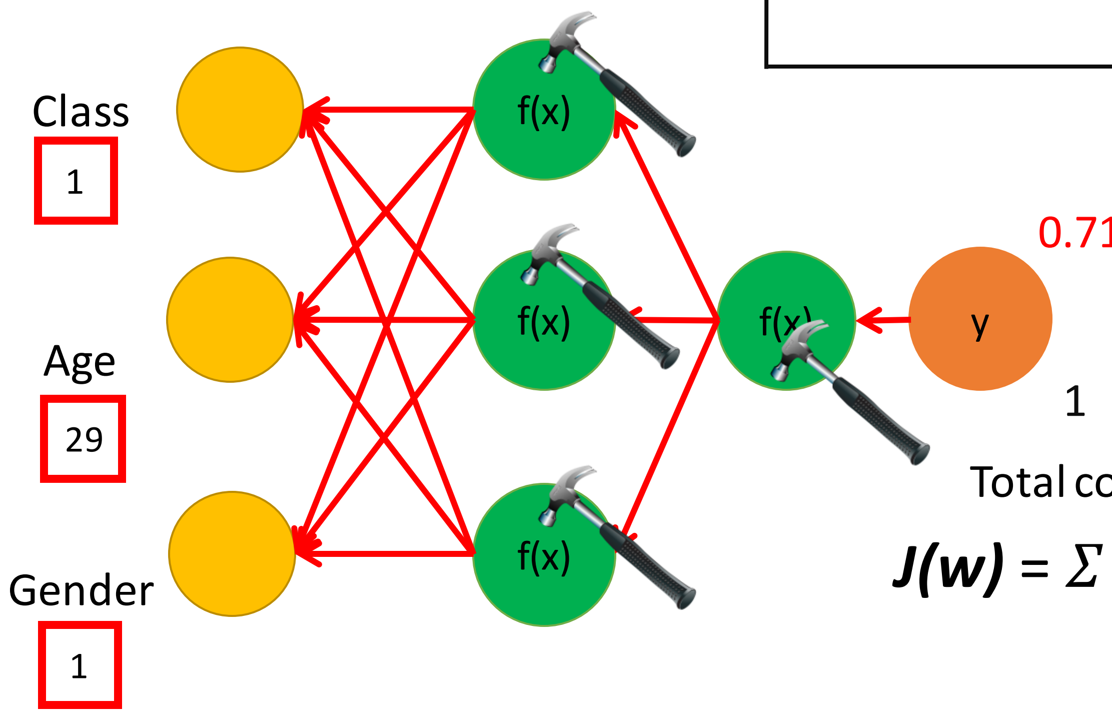
Total cost:

$$J(w) = \Sigma\, (y' - y)^2$$

Neural Network

Passenger Class	Gender	age	Survived
1	1	29	1
1	2	2	0
2	2	21	1
2	1	19	1

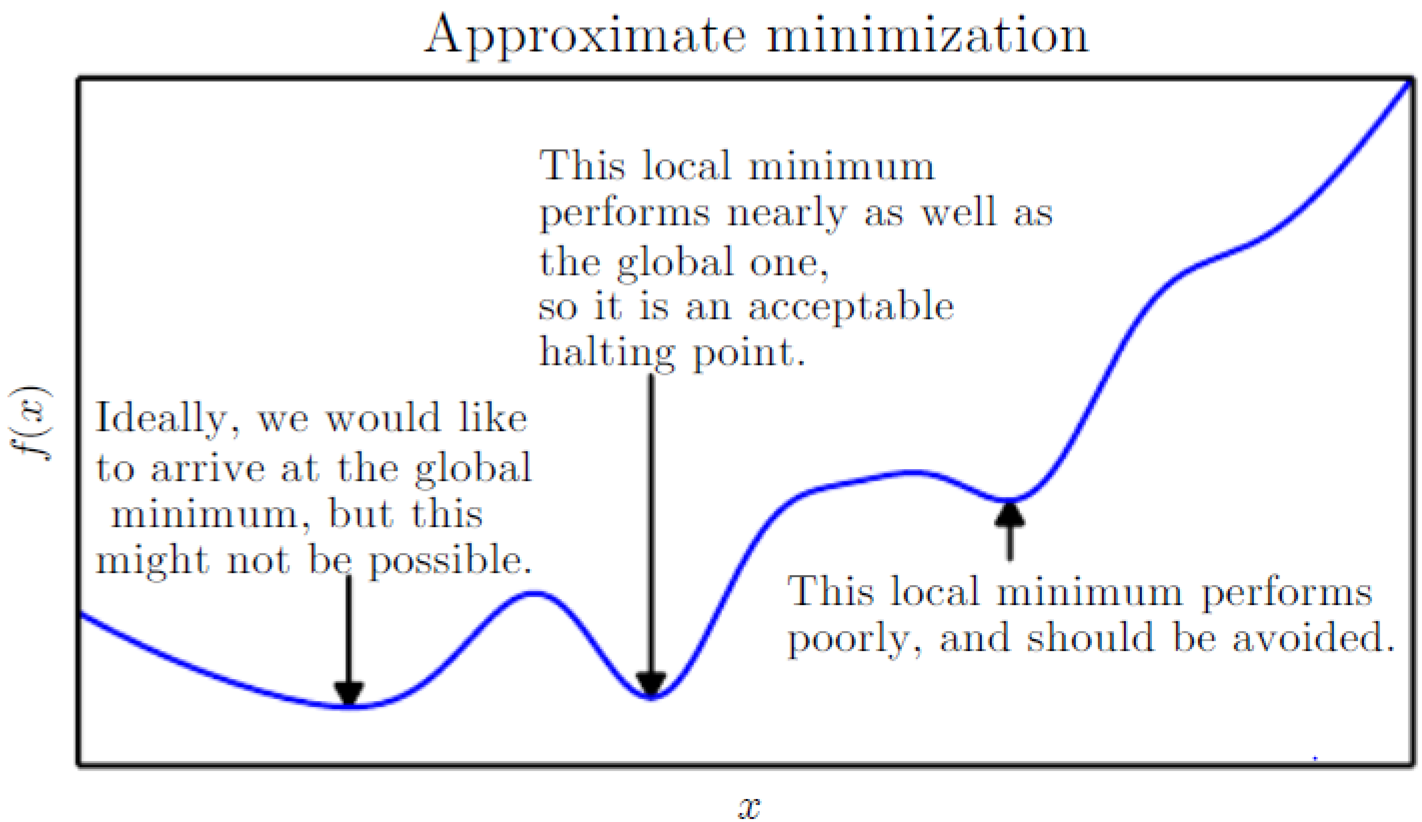




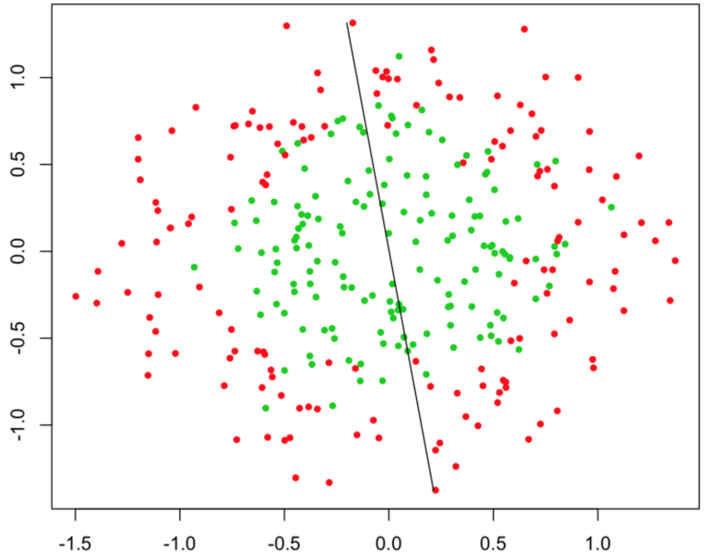
Total cost:

$$J(w) = \Sigma\ (y' - y)^2$$

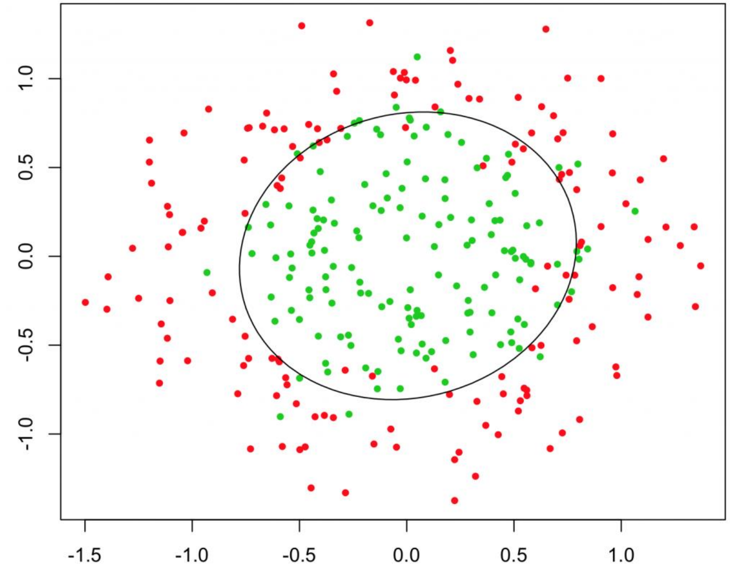
Neural Network



Neural Network cont'd

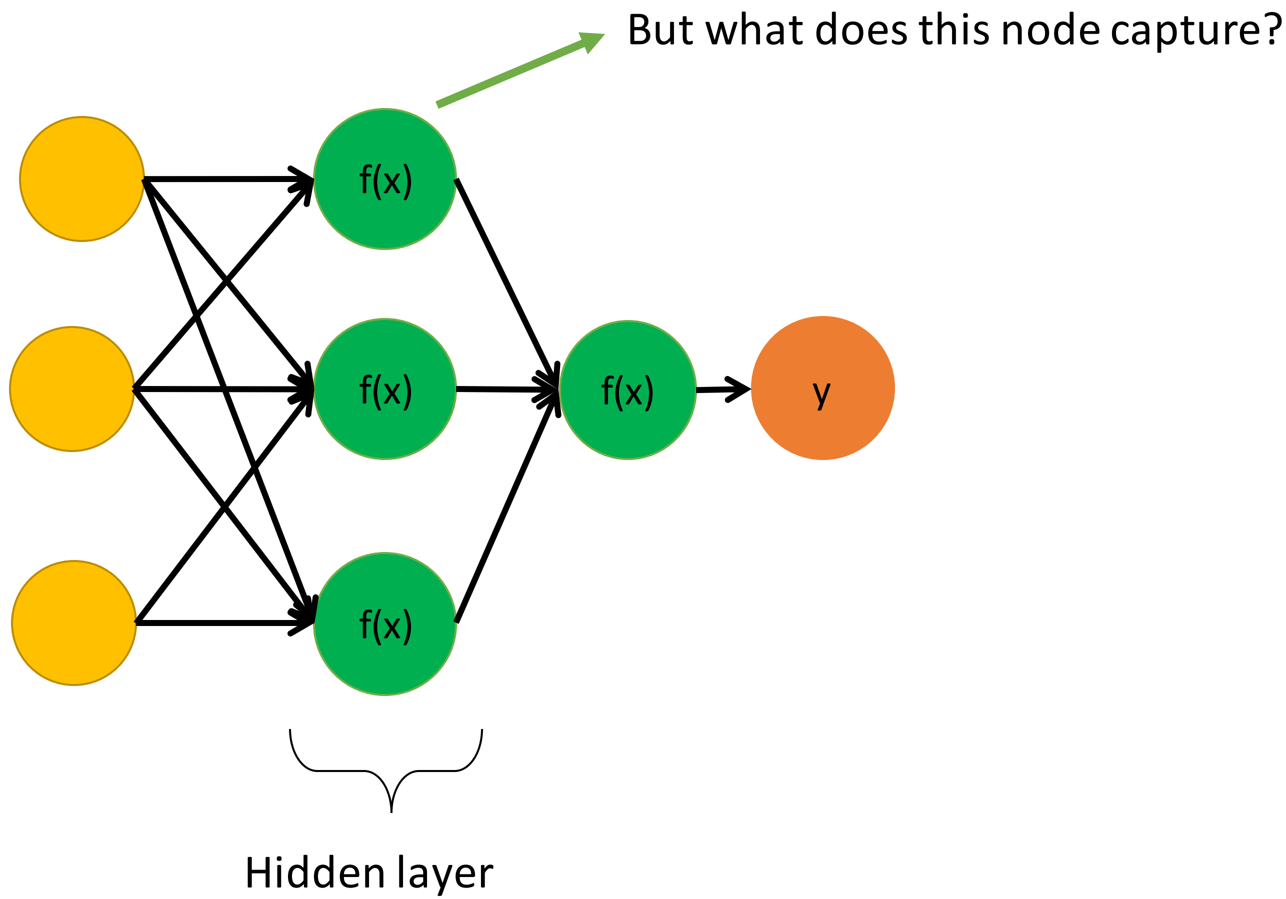


Decision boundary made by a linear model (logistic regression)



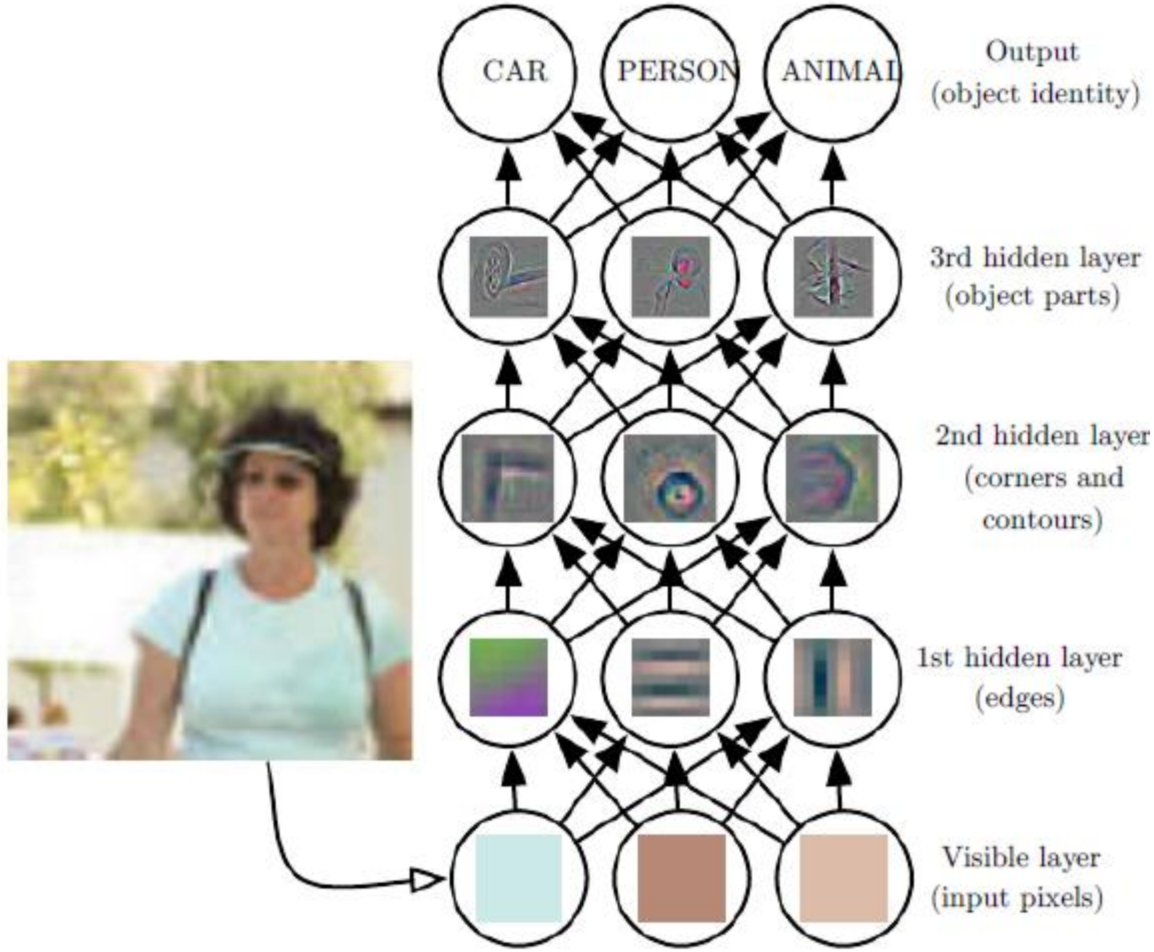
Decision boundary made by a neural network

Neural Network cont'd



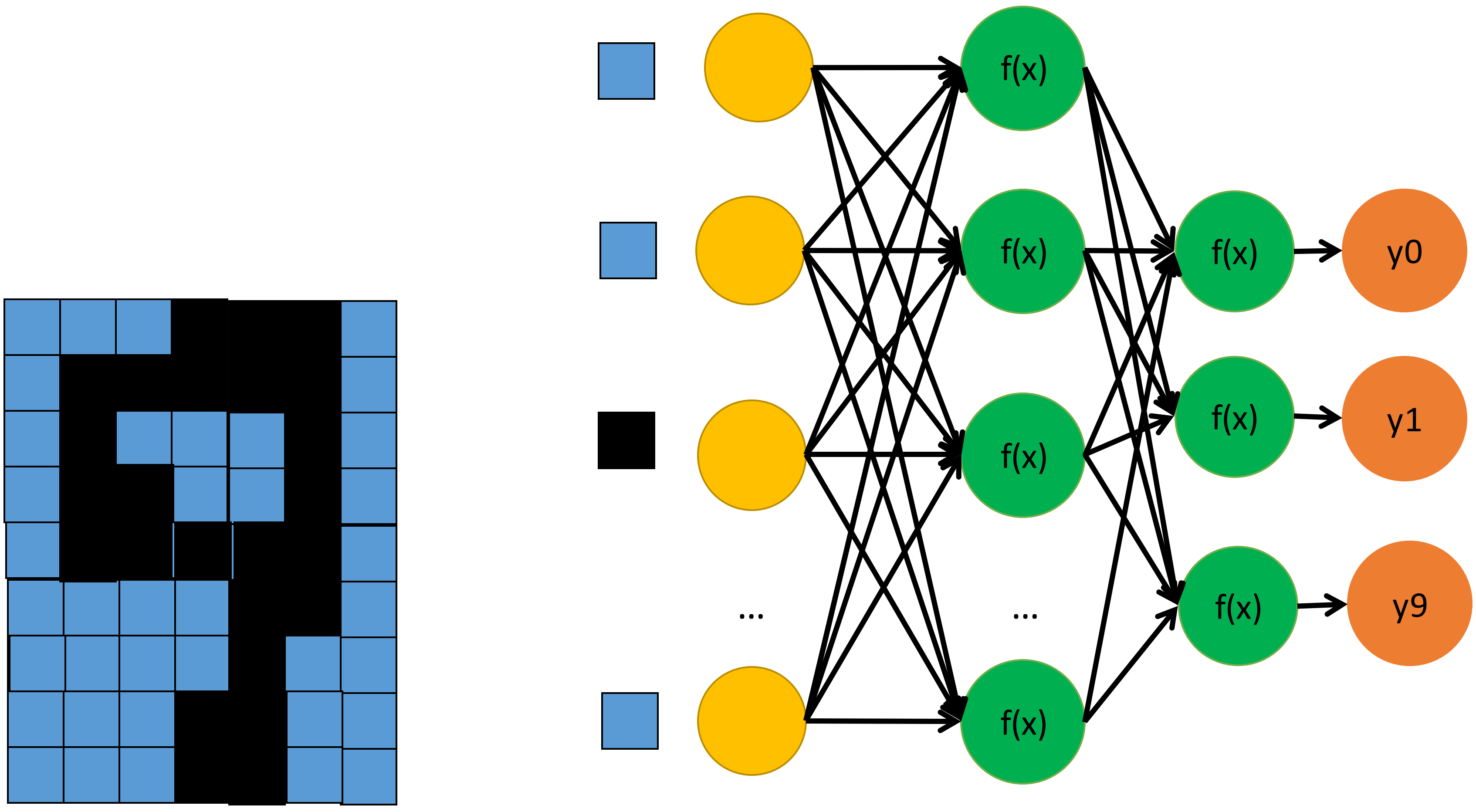
Neural Network cont’d

successive layers can learn higher-level features

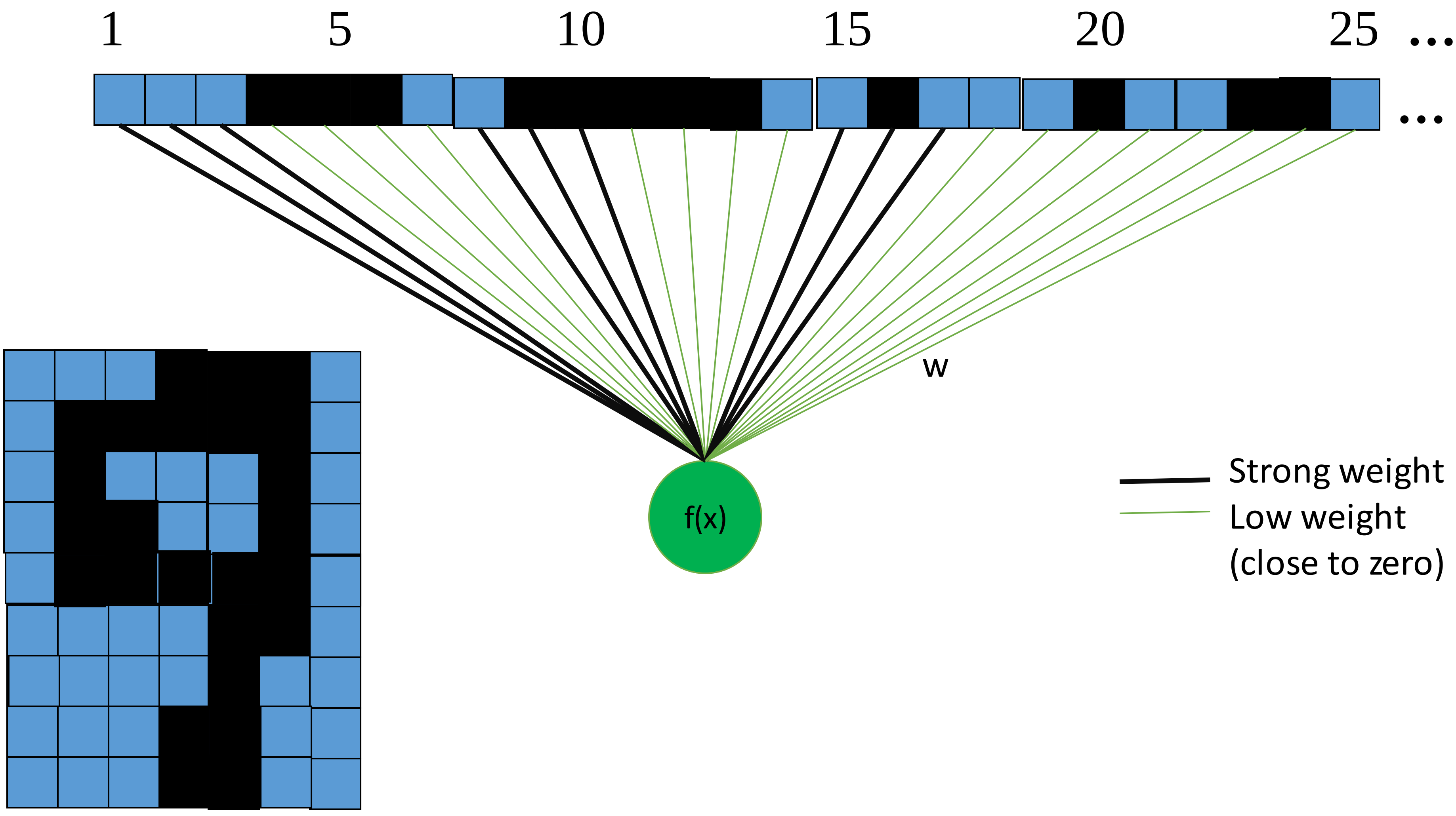


Neural Network cont'd

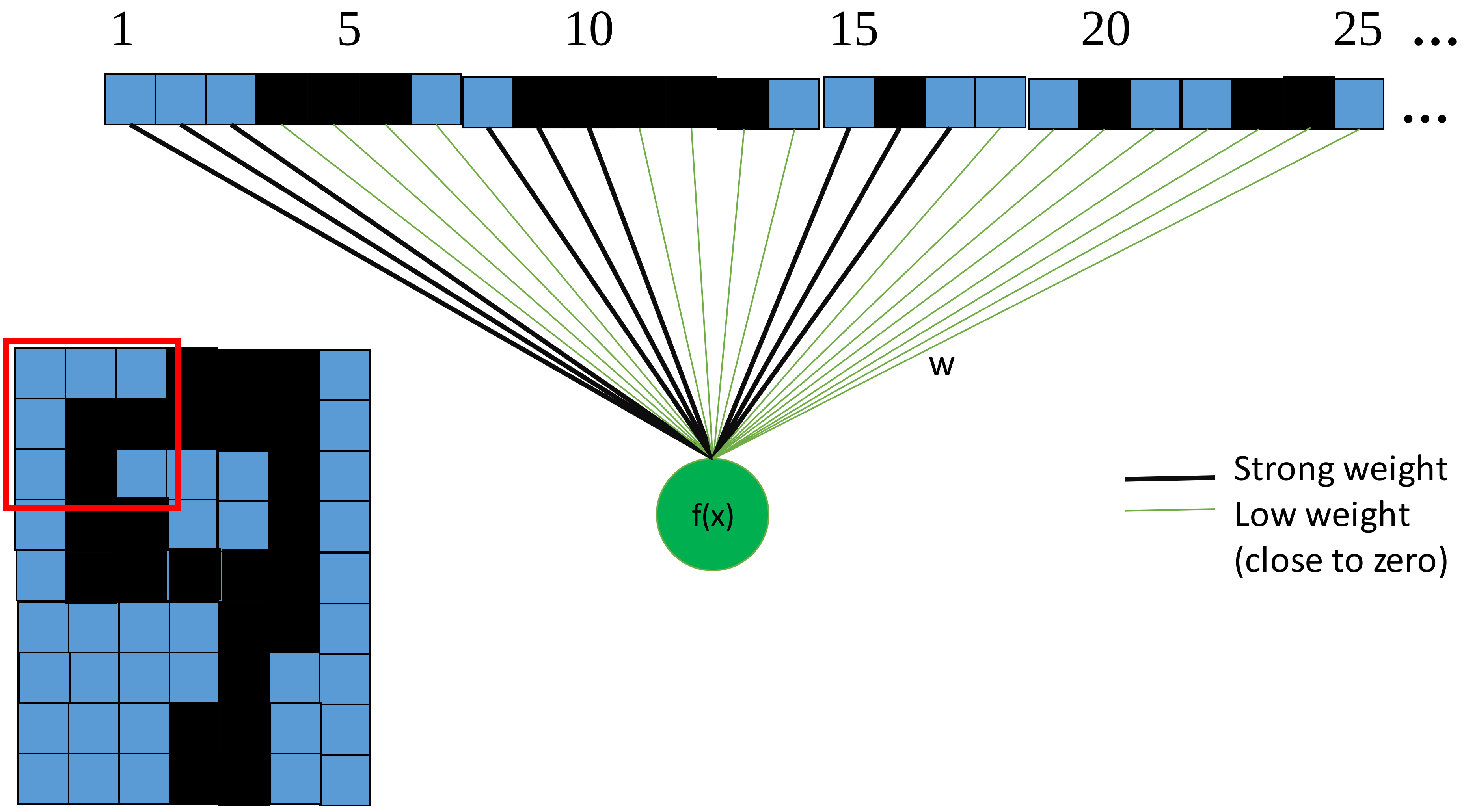
Neural Network cont'd



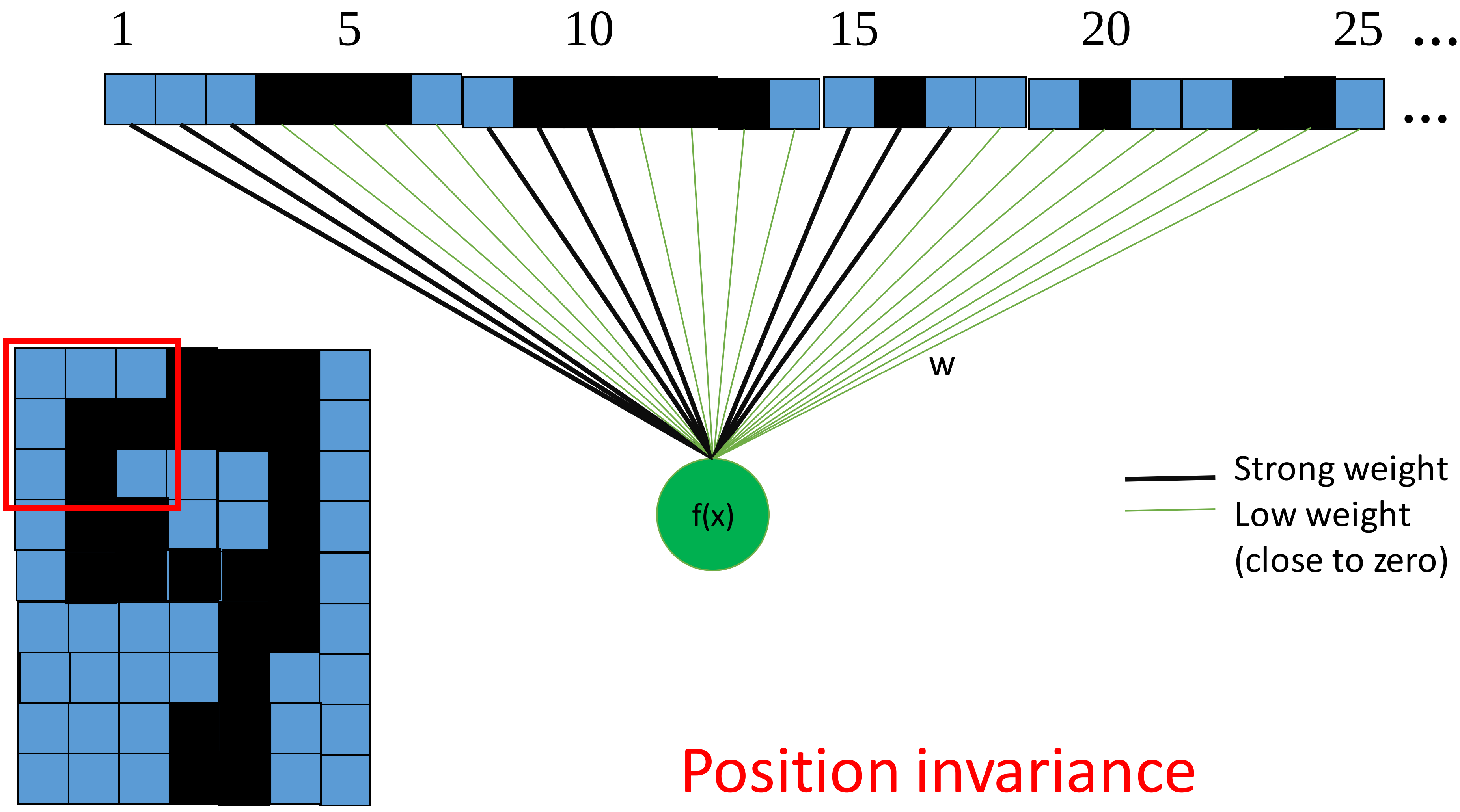
Neural Network cont’d



Neural Network cont’d

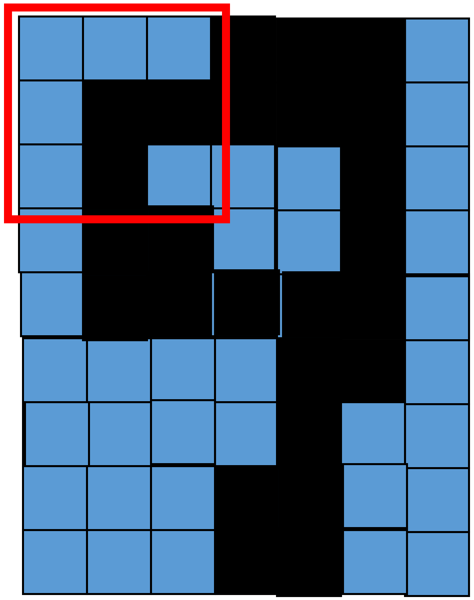


Neural Network cont'd

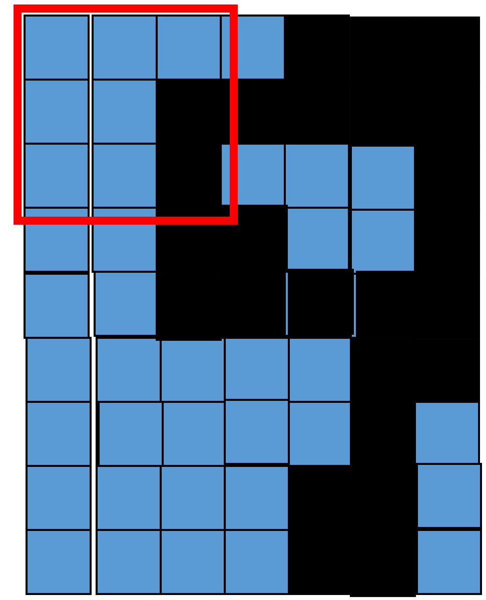


Position invariance

Neural Network cont'd



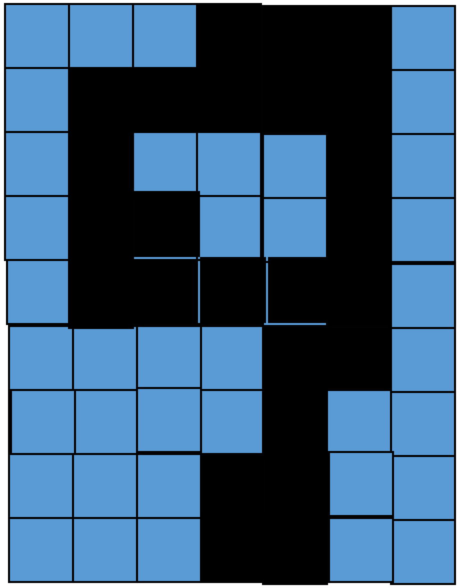
A training sample.



A testing sample. The old hidden unit failed to find the curve.

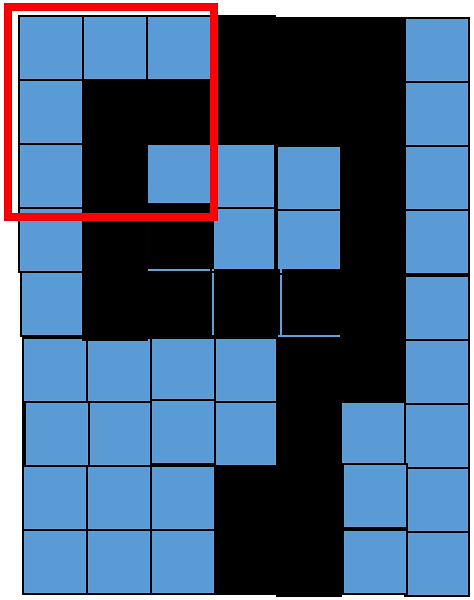
Position invariance

Convolutional Neural Network



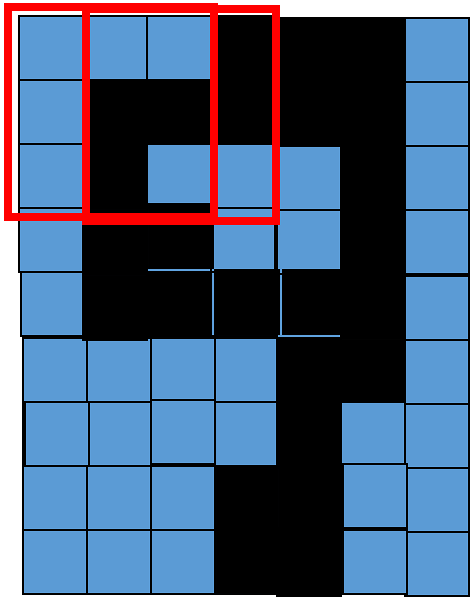
A training sample.

Convolutional Neural Network



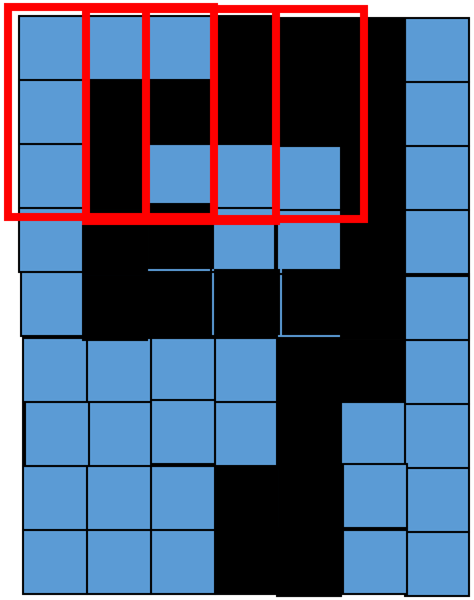
A training sample.

Convolutional Neural Network



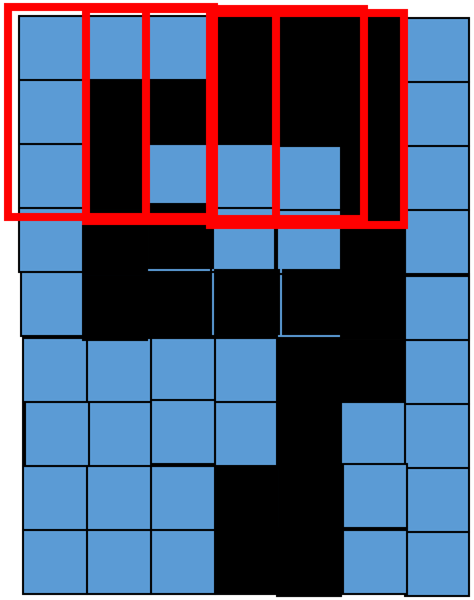
A training sample.

Convolutional Neural Network



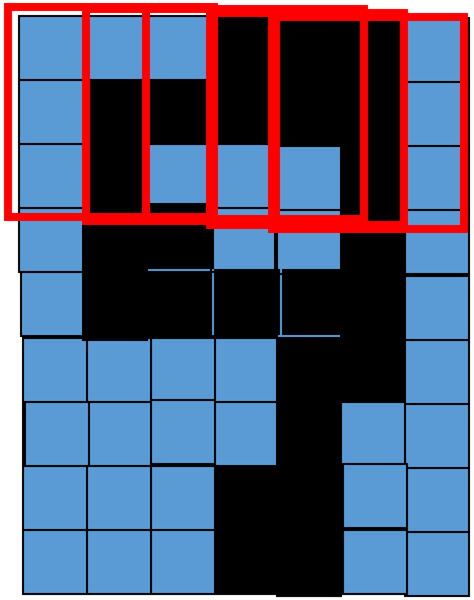
A training sample.

Convolutional Neural Network



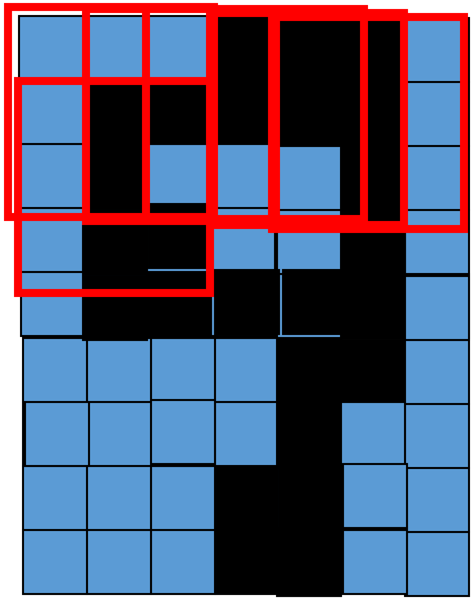
A training sample.

Convolutional Neural Network



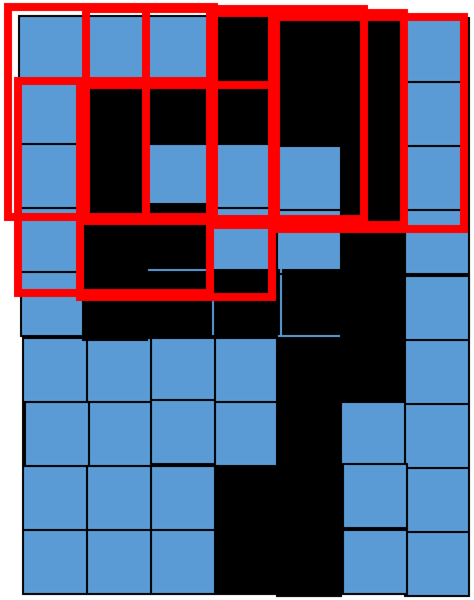
A training sample.

Convolutional Neural Network



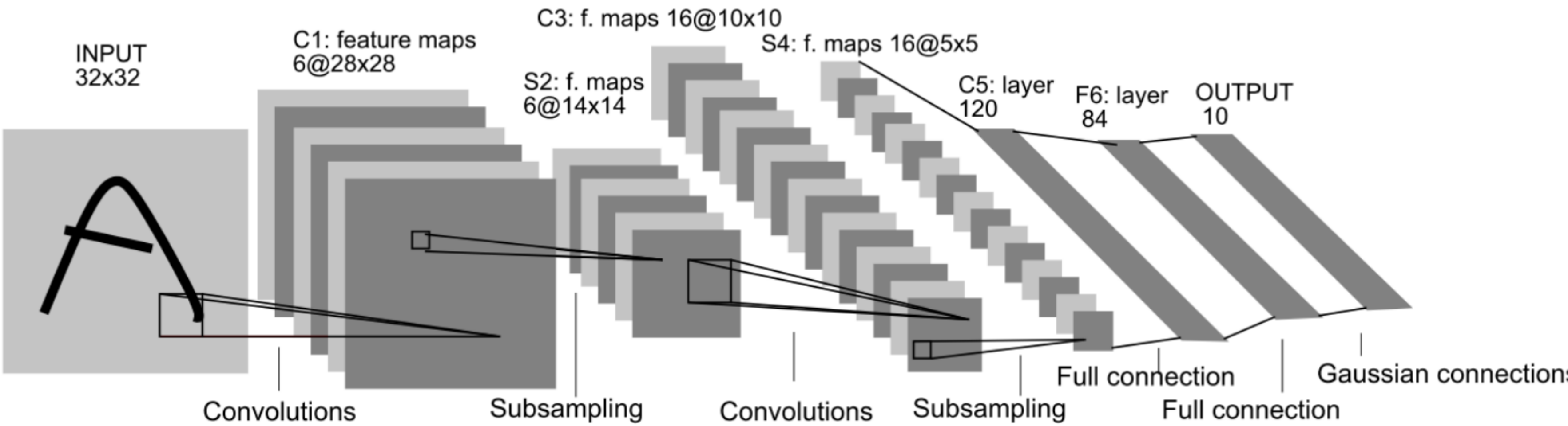
A training sample.

Convolutional Neural Network



A training sample.

Convolutional Neural Network

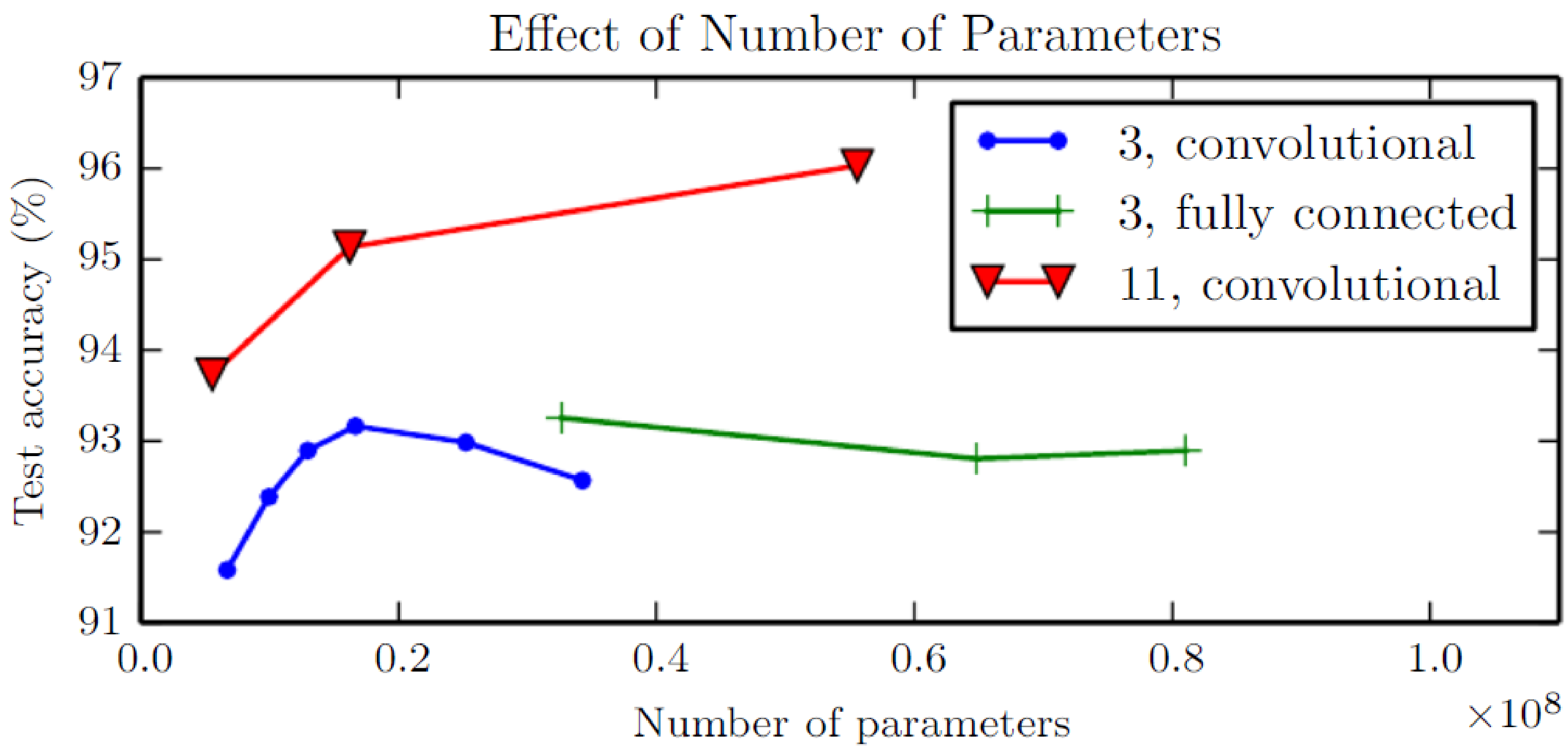


LeNet 5, Layer C5

Why going deep?

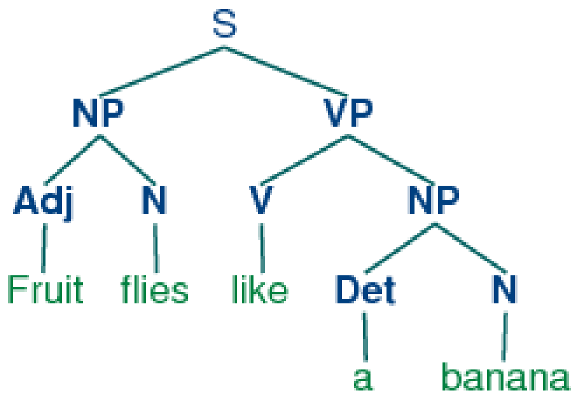
- Most deep learning models have been around for more than 25 years.
    - Theoretically even a NN with a single hidden layer can fit any function.
    - Gradient vanishing problem solved by new set of training algorithms and models.
- Deep structure:
    - Successive layers can learn higher-level features (latent factors)
    - Replace manual feature engineering.
    - Better generalizability. Can represent more complex functions with less parameters.

Why going deep?



What else?

- Classification
- Regression
- Sequential output
- Structured output



Thank you!