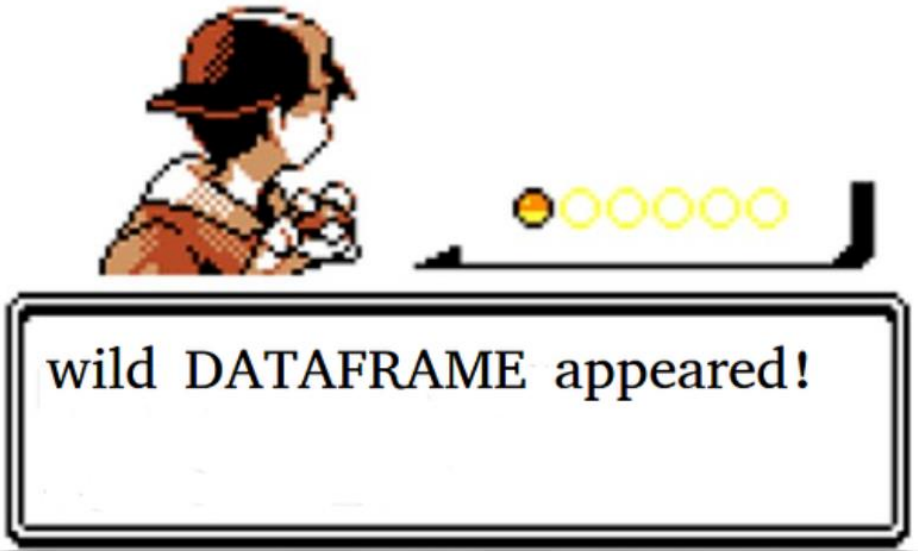
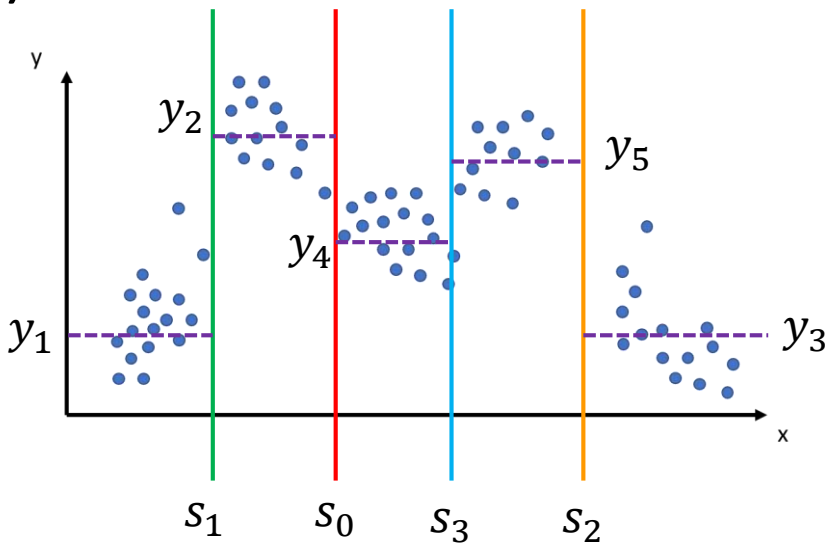
CISC 372
Advanced Data Analytics
L10 AutoDiff4DL

	name	age	state	num_children	num_pets
0	john	23	iowa	2	0
1	mary	78	dc	2	4
2	peter	22	california	0	0
3	jeff	19	texas	1	5
4	bill	45	washington	2	0
5	lisa	33	dc	1	0



Tree[s]

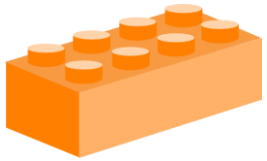
- Tree Induction
- Information Gain/Gain Ratio/Gini Index
- ID3, CART, C4.5
- Splitting Numeric Attribute
- Feature Selection (is difficult)
- Random Forest (the easy way)
  - Built-in bootstrap sampling
- Regression Tree
- XGBoost



Today

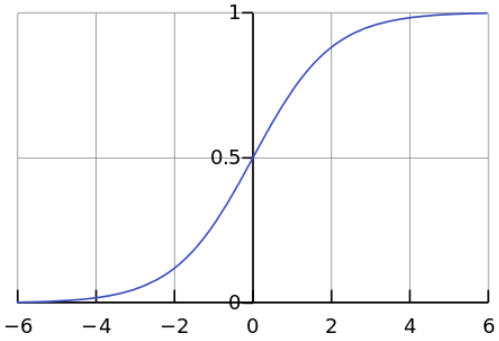
- AutoDiff
- Neural Network
- Convolutional Neural Network

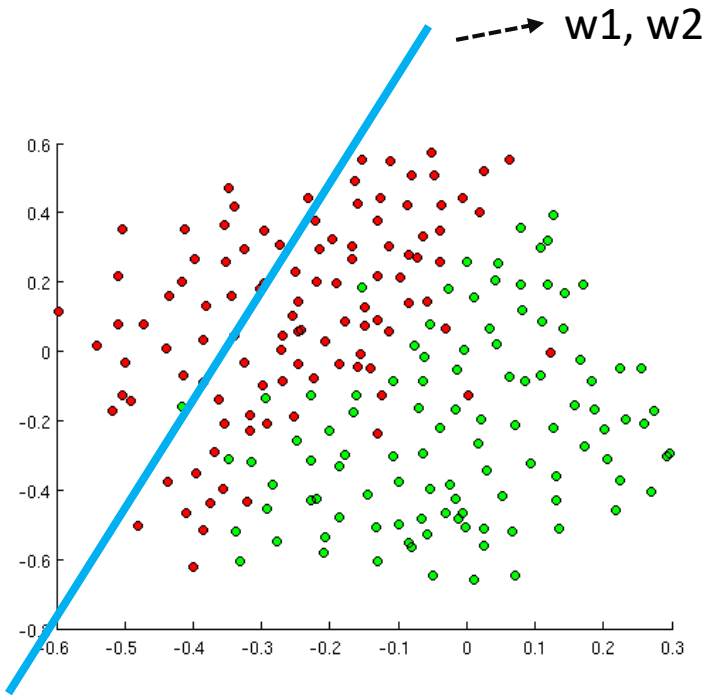
Logistic Regression

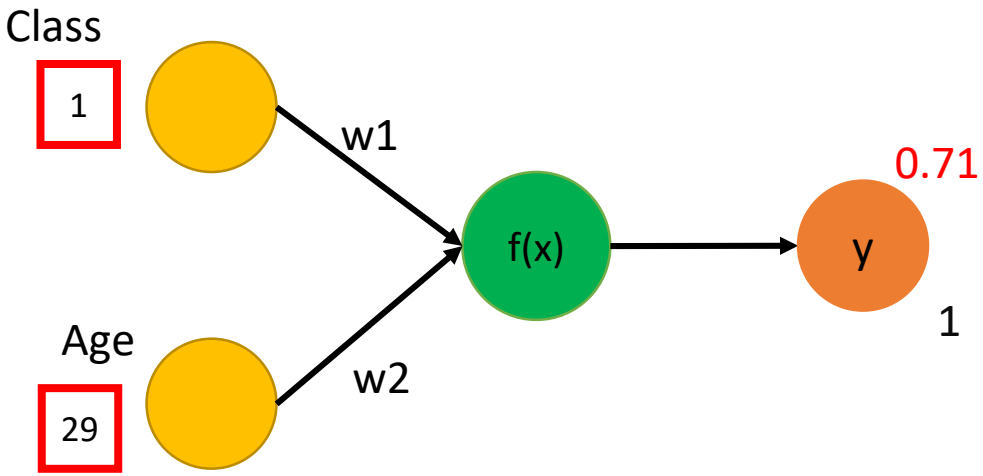


Passenger Class	Age	Survived
1	29	1
1	2	0
2	21	1
2	19	1
…		

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







a = w1 * Class + w2 * Age

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

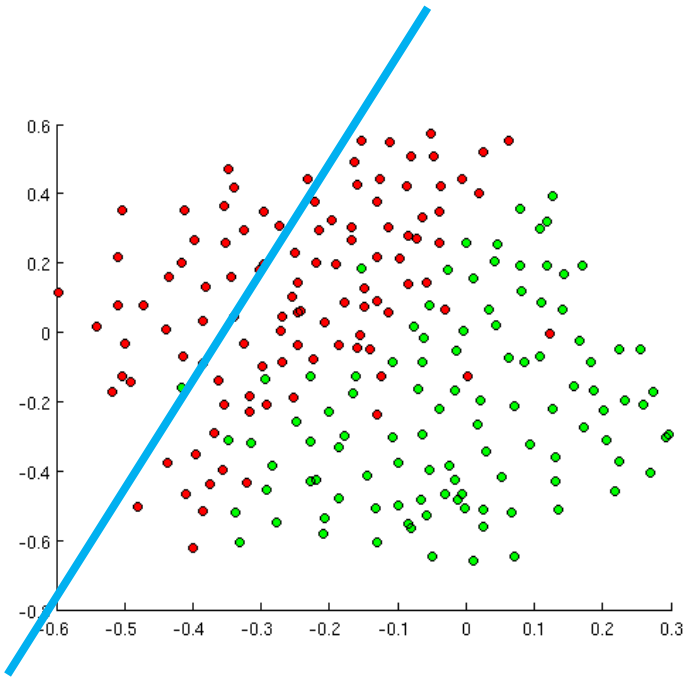
y = f(a)

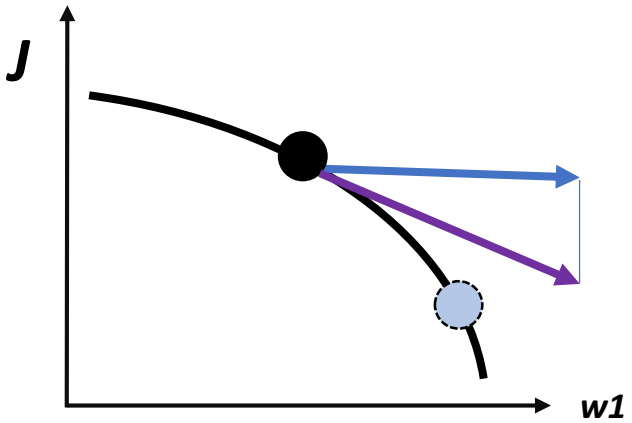
y' = f(0.9) = 0.71

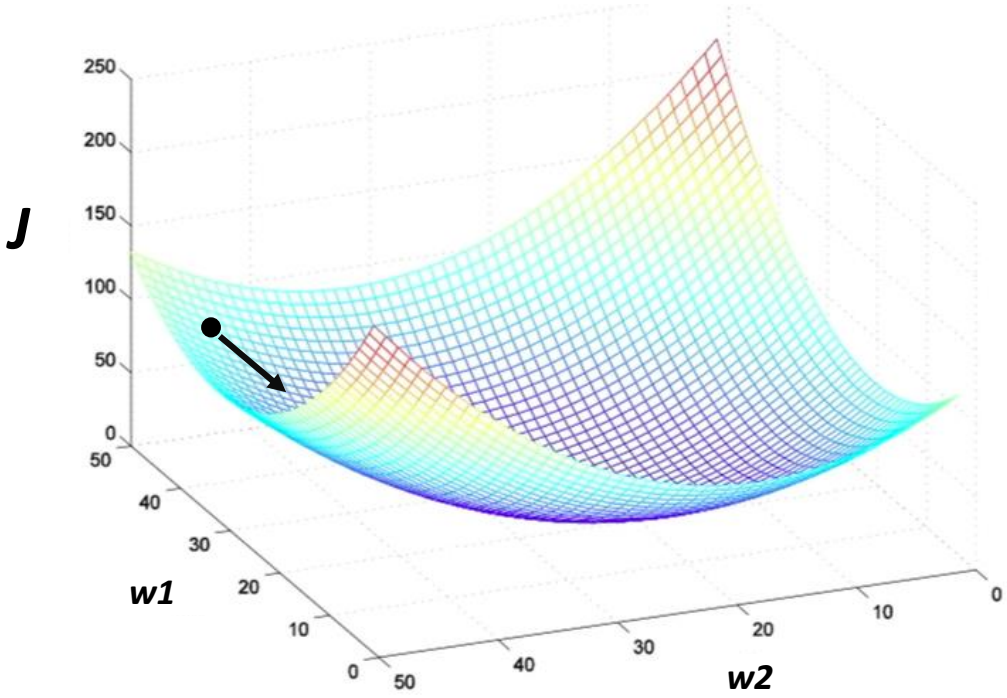
Total cost:

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

Logistic Regression cont'd



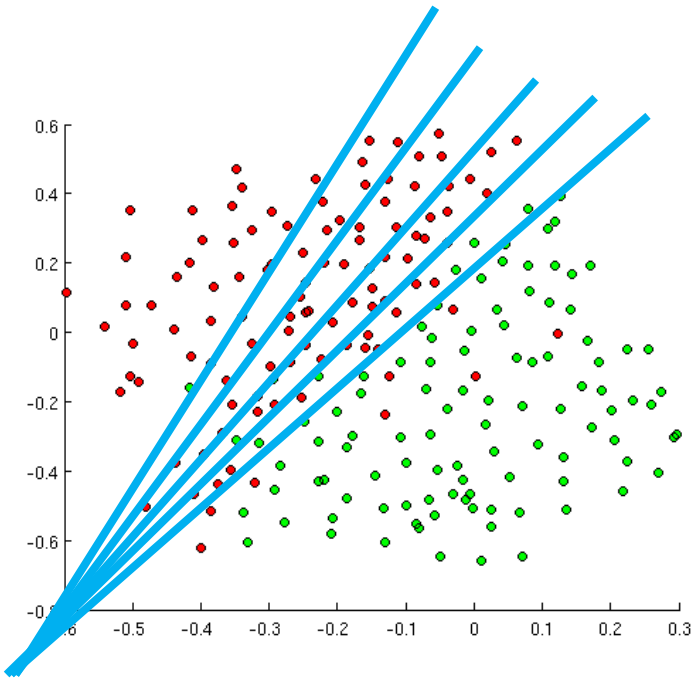


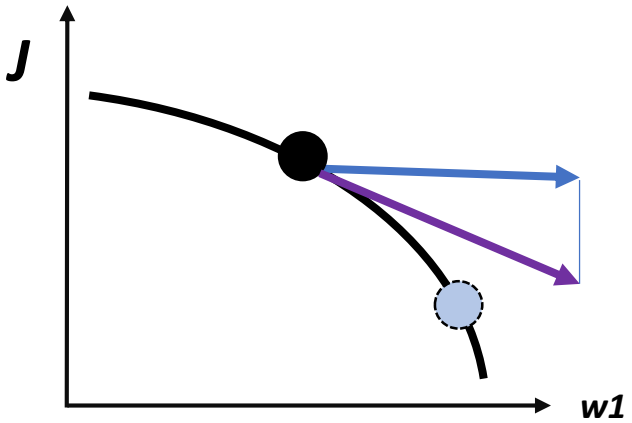


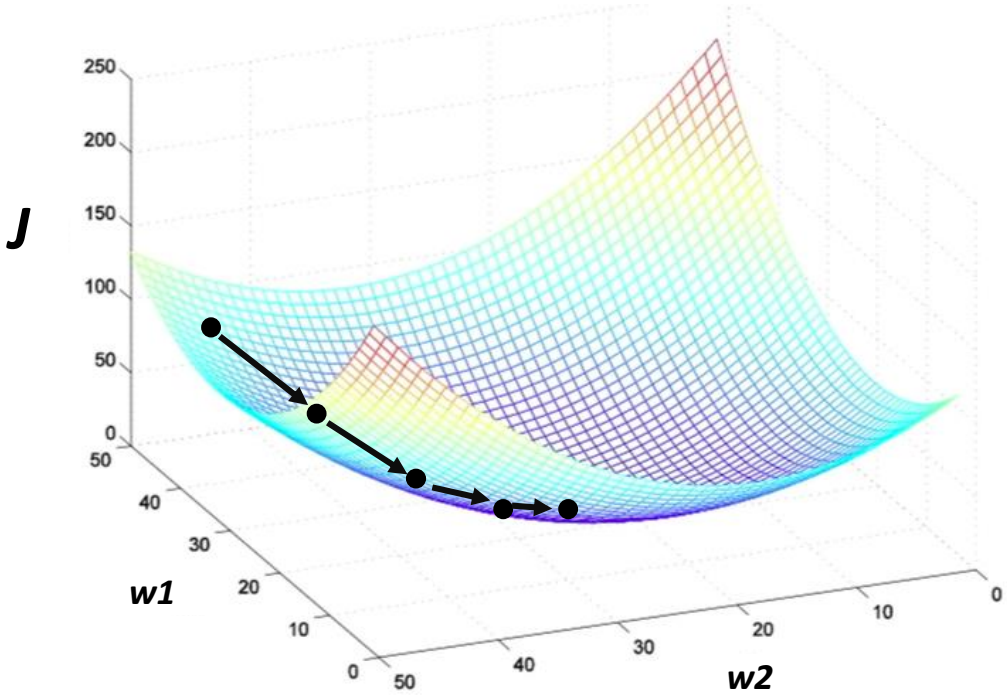
Total cost:

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

Logistic Regression cont'd



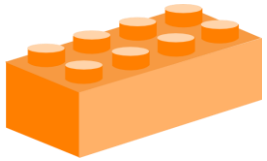


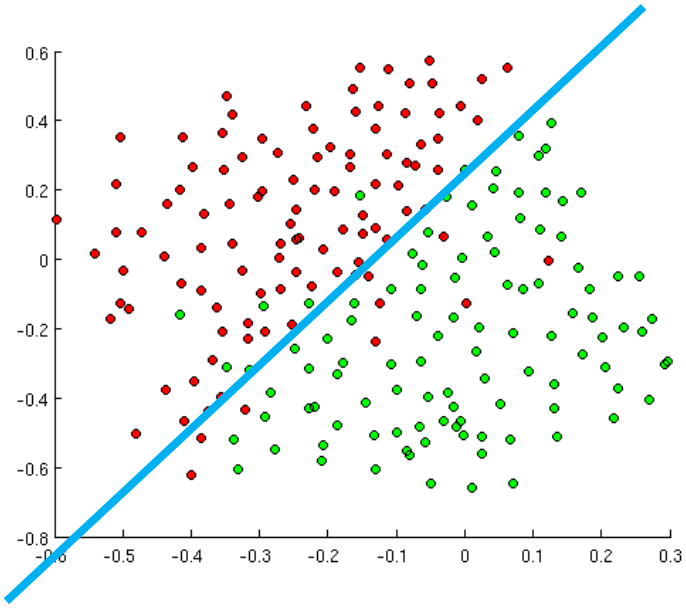


Total cost:

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

Logistic Regression cont'd





Decision boundary made by a linear model (logistic regression)

Nonlinearity

- XOR?
  - Not linearly separable
    - Convexity (convex set)
- Feature Map
  - Aka basis function, kernels etc
  - a,b -> class
  - a,b,ab -> class
  - Now it is linearly separable
  - Difficult to find the right function

	a	b	class
1	0	0	0
2	0	1	1
3	1	0	1
4	1	1	0

	a	b	ab	class
1	0	0	0	0
2	0	1	0	1
3	1	0	0	1
4	1	1	1	0

Nonlinearity

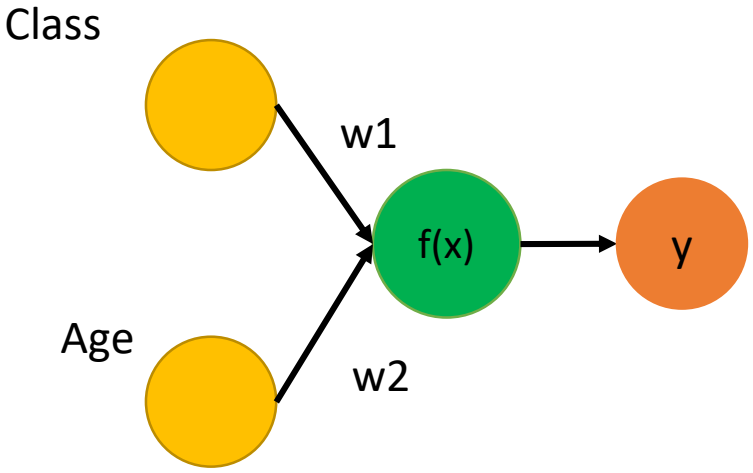
- XOR?
  - Not linearly separable
    - Convexity (convex set)
- Feature Map
  - Aka basis function, kernels etc
  - a,b -> class
  - a,b,ab -> class
  - Now it is linearly separable
  - Difficult to find the right function
- Solution
  - Learn the mapping function from data!
  - Multi-layer perception

	a	b	class
1	0	0	0
2	0	1	1
3	1	0	1
4	1	1	0

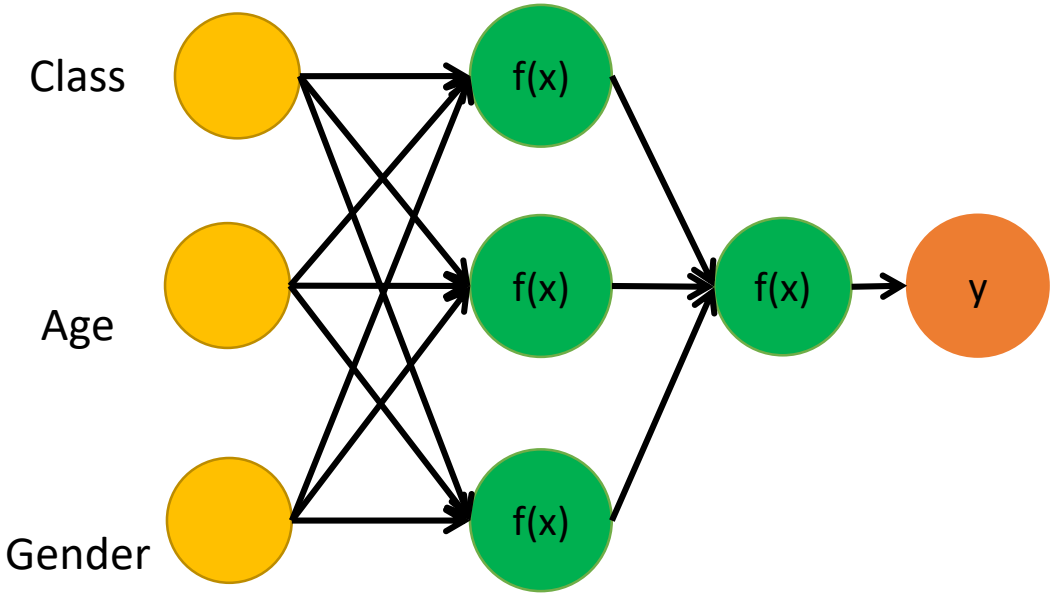
	a	b	ab	class
1	0	0	0	0
2	0	1	0	1
3	1	0	0	1
4	1	1	1	0

Neural Network

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



A logistic regression model.

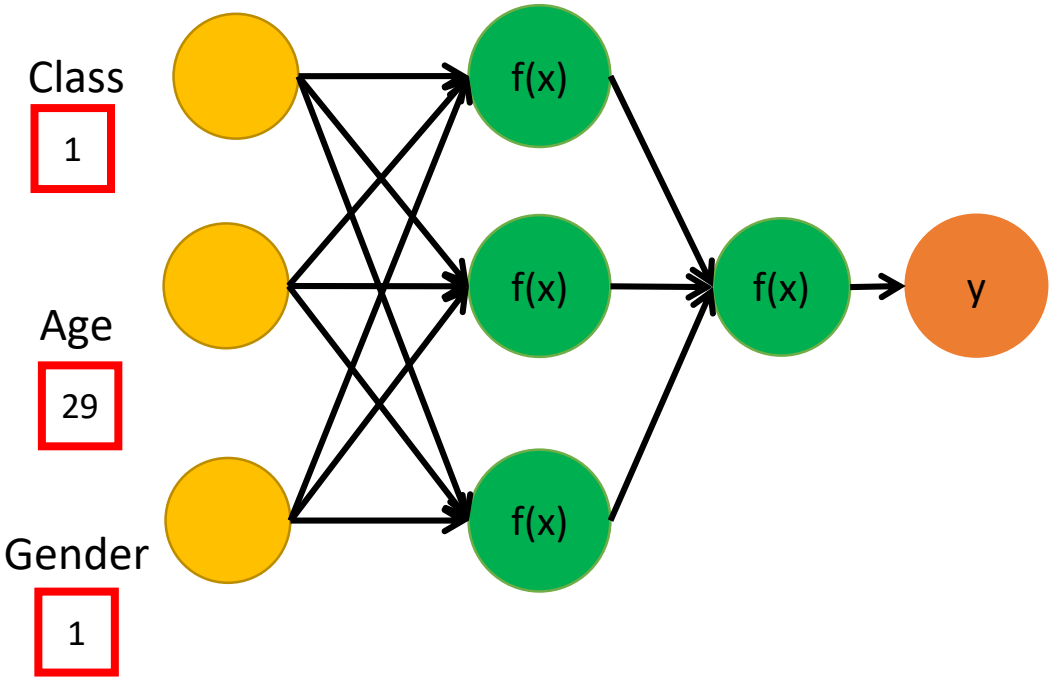


A neural network model.

Neural Network

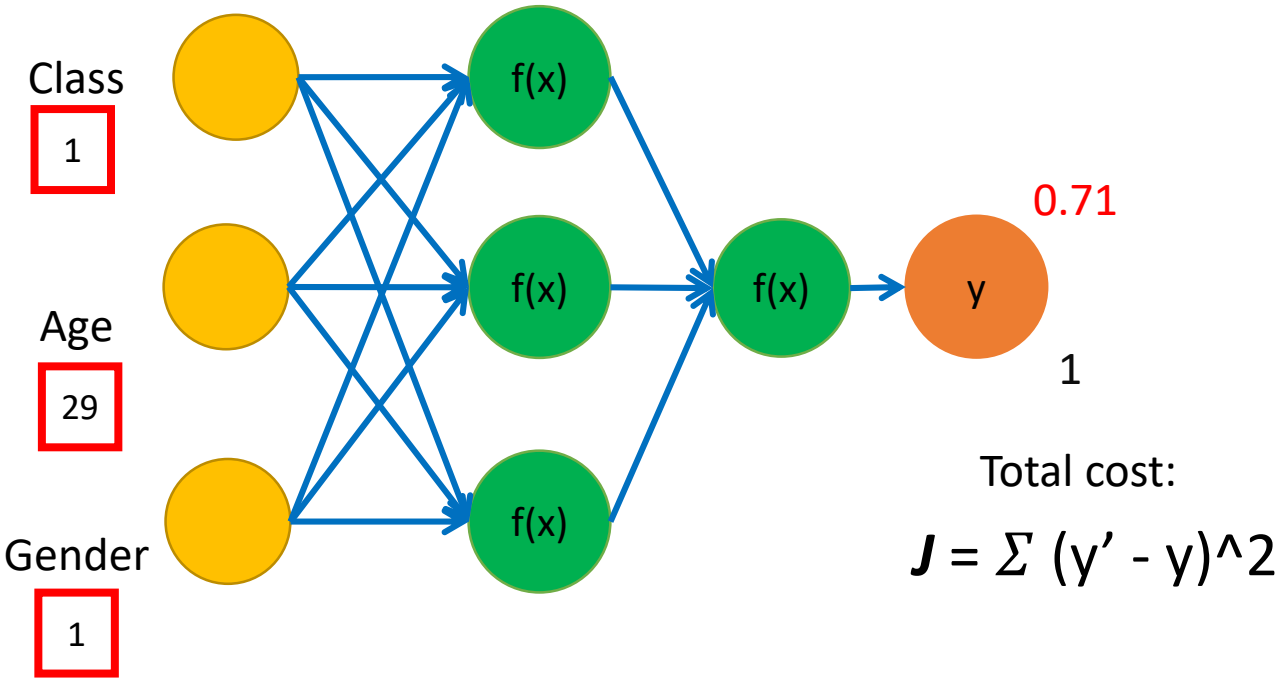
Passenger Class	Gender	age	Survived
1	1	29	1
1	2	2	0
2	2	21	1
2	1	19	1

Randomly initialize weights.



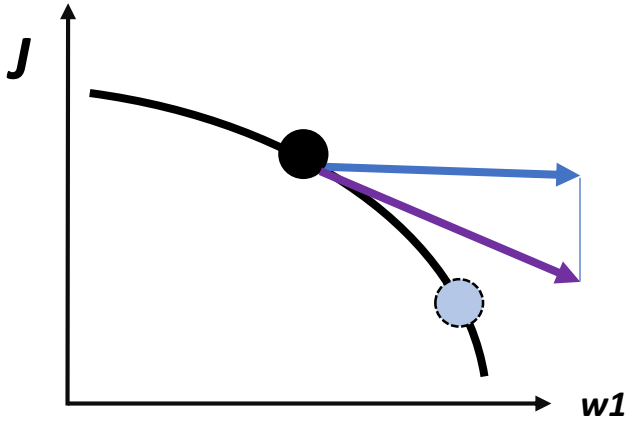
Neural Network

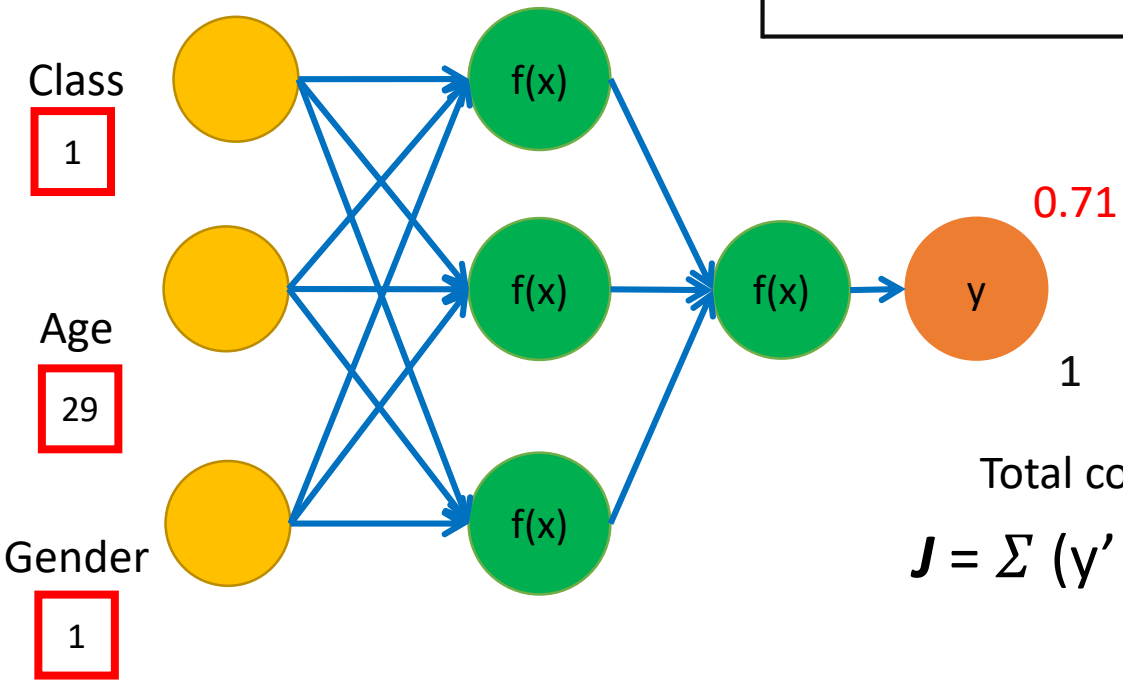
Passenger Class	Gender	age	Survived
1	1	29	1
1	2	2	0
2	2	21	1
2	1	19	1



Neural Network

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



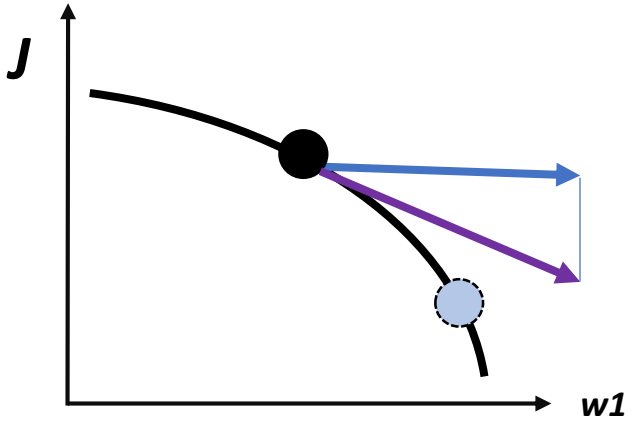


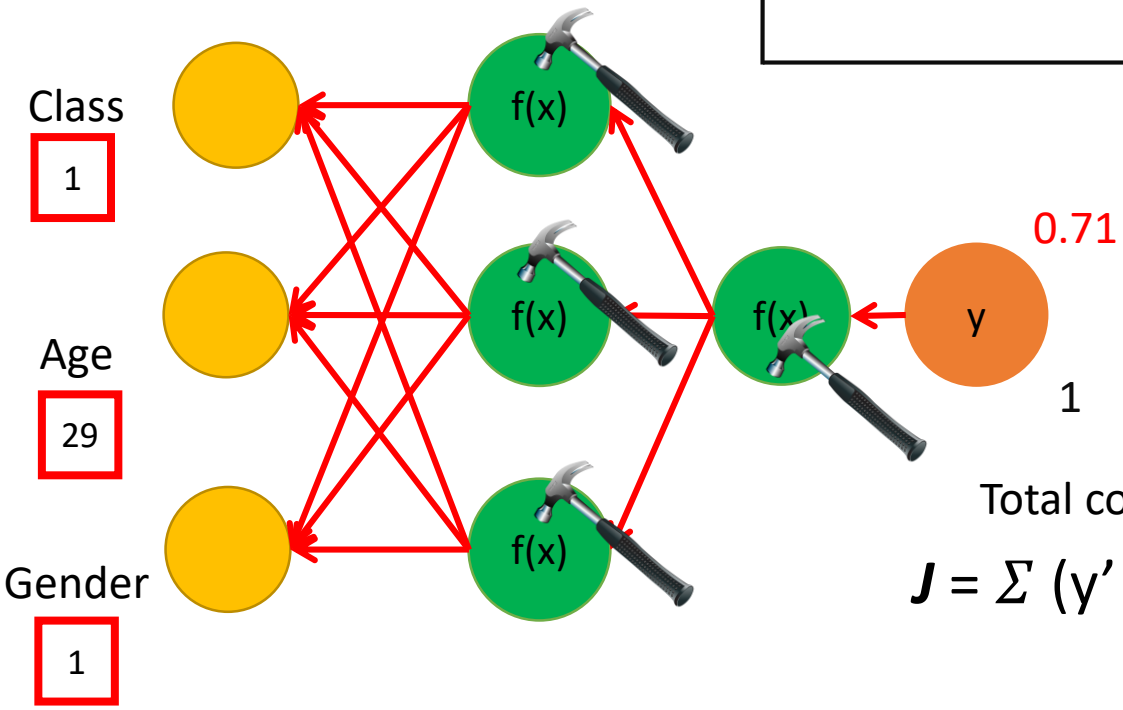
Total cost:

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

Neural Network

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





Automatically Differentiation engine

- Chain Rule:

$$z = \omega x + b$$
$$\hat{y} = \sigma(z)$$
$$L = \frac{1}{2}(y - \hat{y})^2 + \lambda\frac{1}{2}\omega^2$$

$$L = \frac{1}{2}(y - \sigma(\omega x + b))^2 + \lambda\frac{1}{2}w^2$$

$$\frac{\partial L}{\partial_\omega} = \frac{\partial}{\partial\omega}\left(\frac{1}{2}(y - \sigma(\omega x + b))^2 + \lambda\frac{1}{2}w^2\right)$$

$$\frac{\partial L}{\partial_\omega} = (\sigma(\omega x + b) - t)\sigma'(wx + b) + \lambda w$$

$$\frac{\partial L}{\partial_b} = \frac{\partial}{\partial w}\left(\frac{1}{2}(y - \sigma(\omega x + b))^2 + \lambda\frac{1}{2}w^2\right)$$

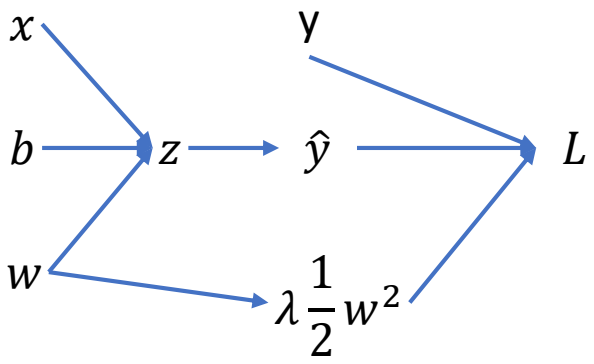
$$\frac{\partial L}{\partial_b} = (6(\omega x + b) - y)\sigma'(\omega x + b)$$

Automatically Differentiation engine

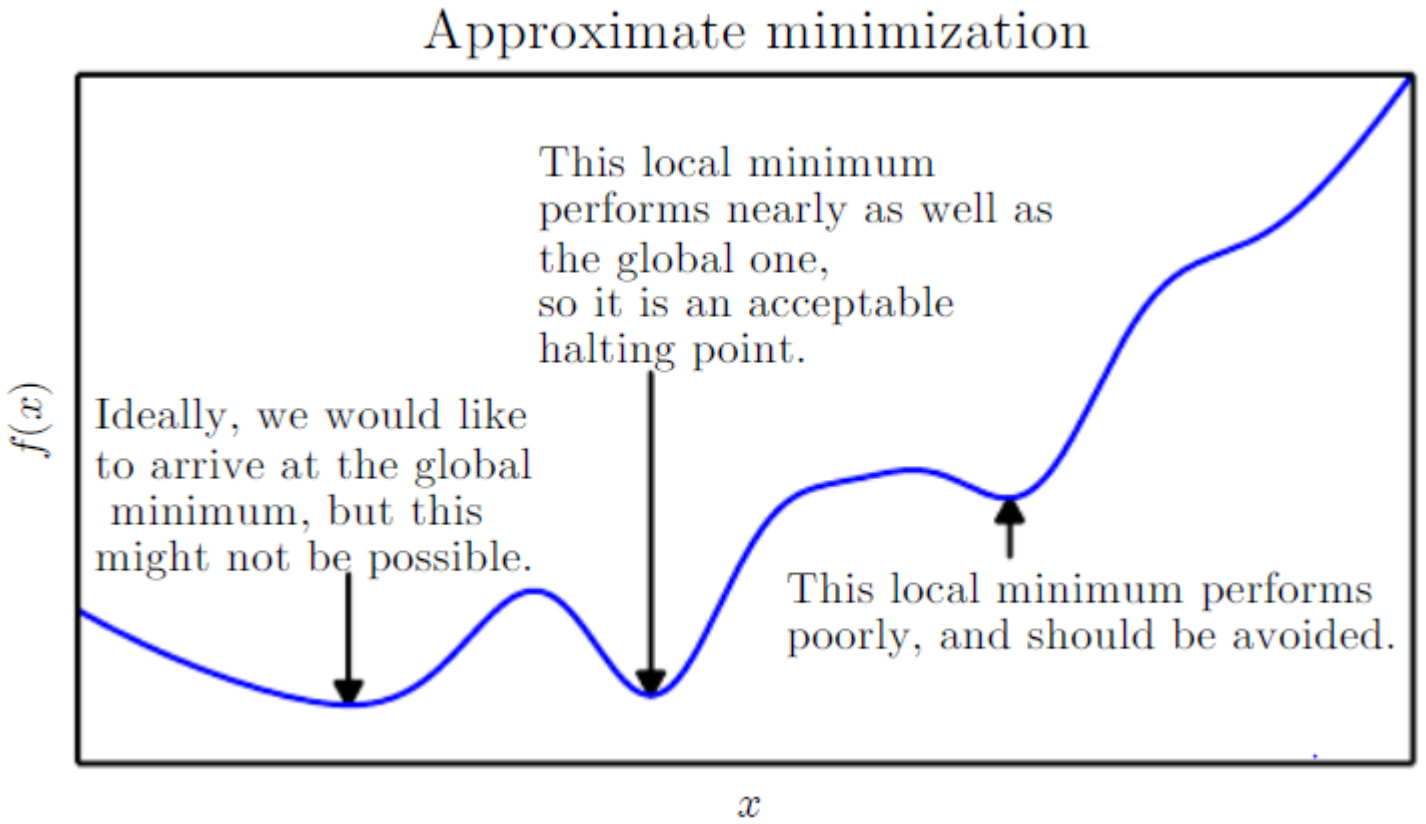
- Chain Rule:

$$\frac{d}{d_t} f\big(x(t), y(t)\big) = \frac{\partial f}{\partial x}\frac{dx}{dt} + \frac{\partial f}{\partial y}\frac{dy}{dk}$$

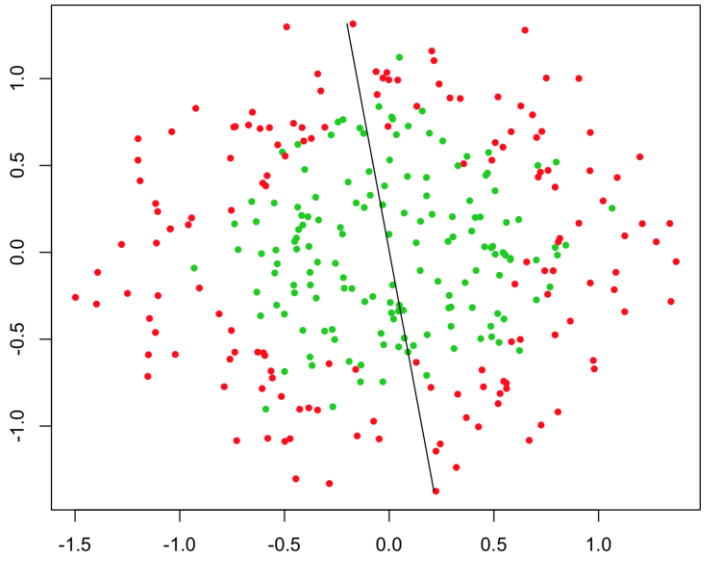
computation graph



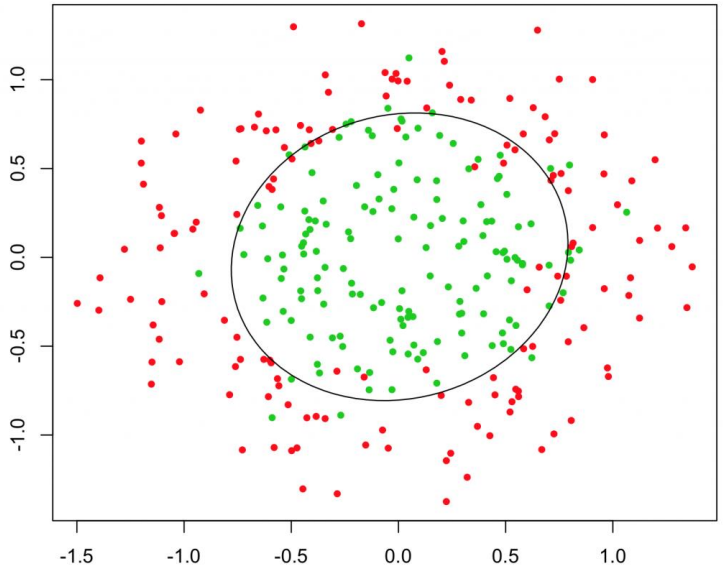
Neural Network



Neural Network cont'd

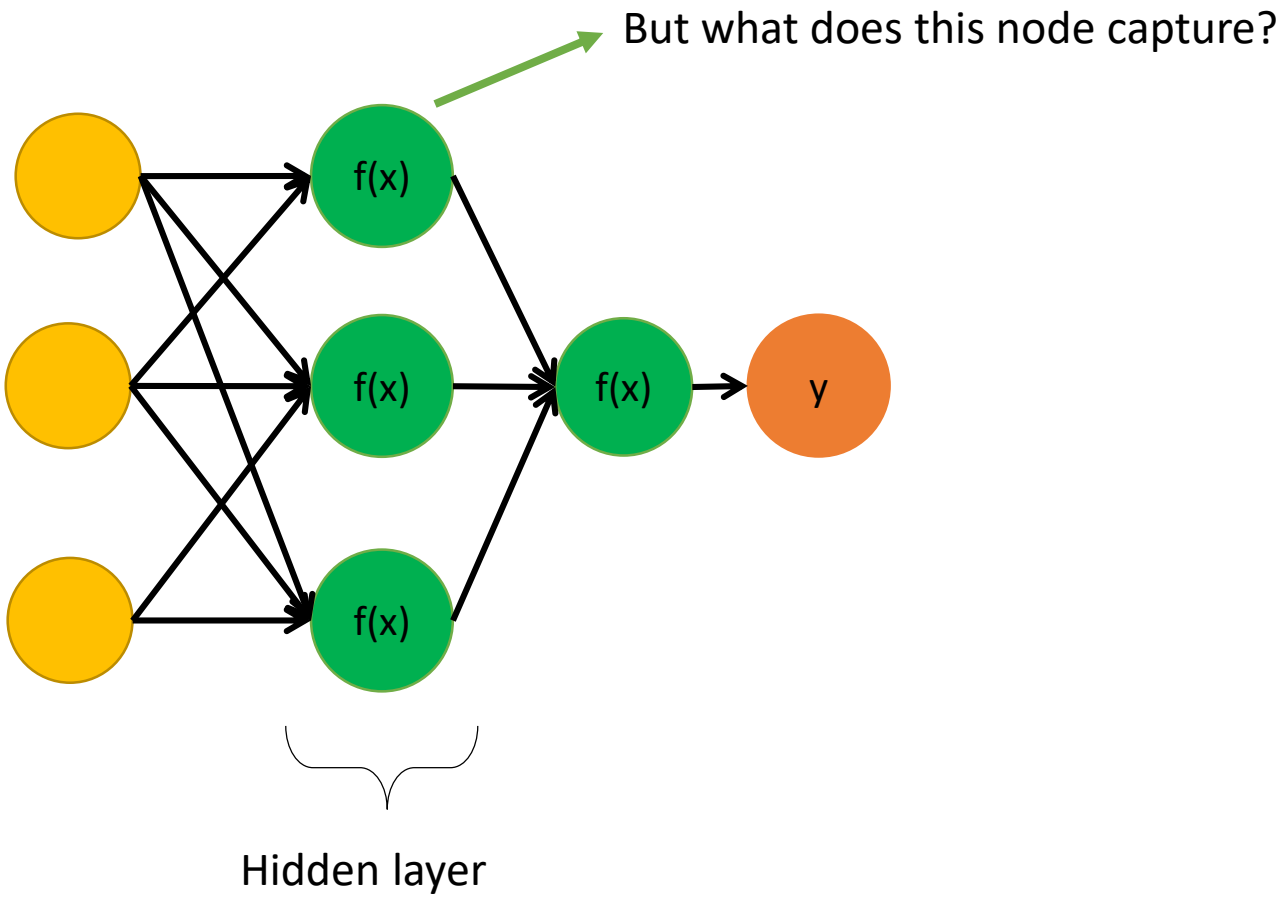


Decision boundary made by a linear model (logistic regression)

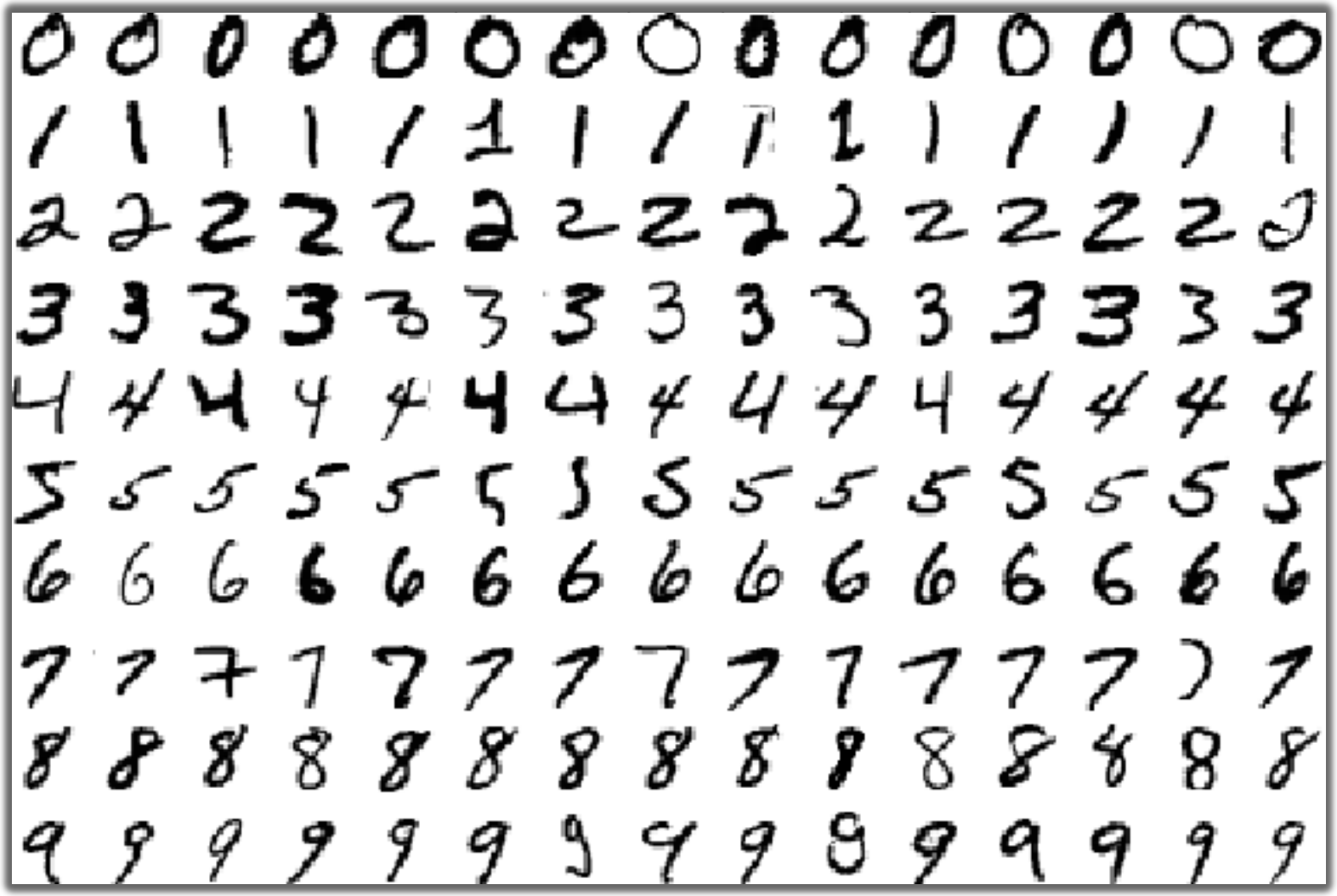


Decision boundary made by a neural network

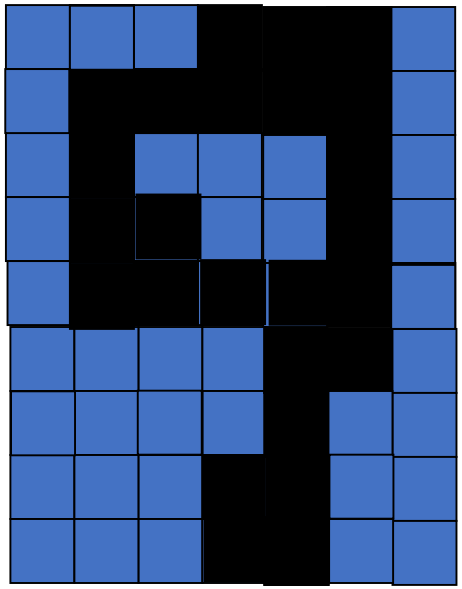
Neural Network cont'd

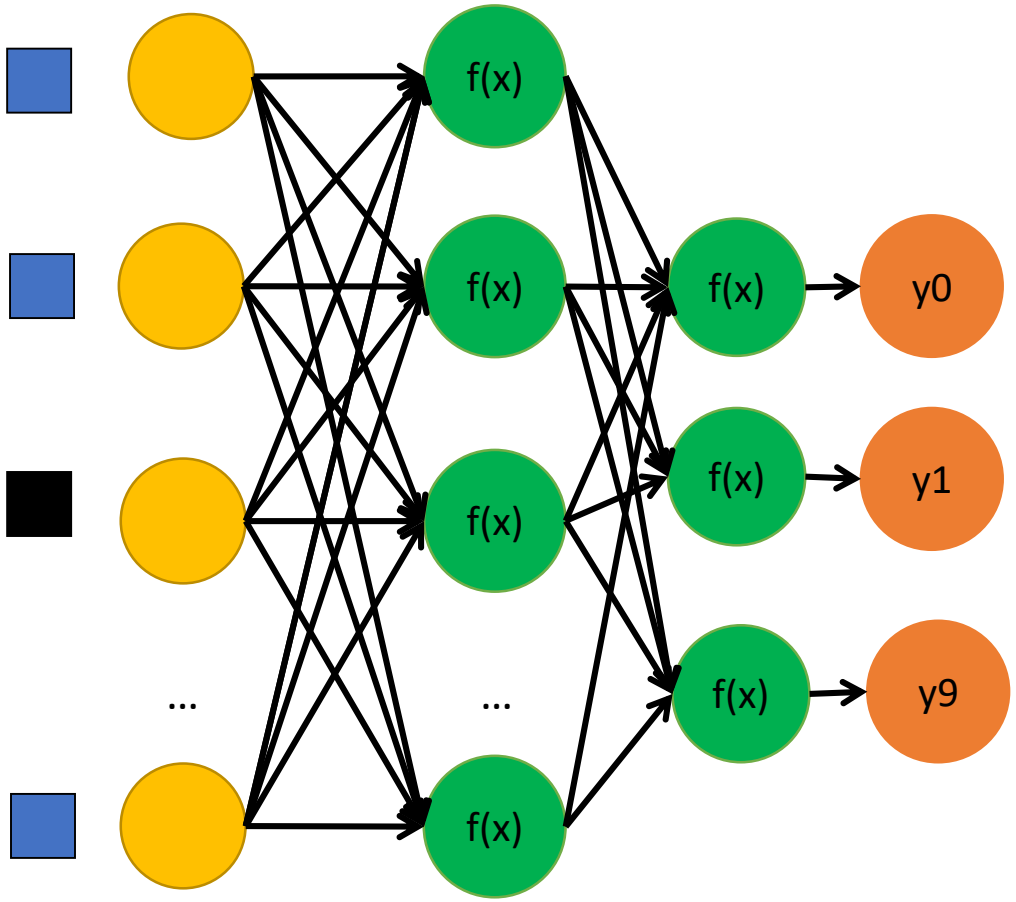


Neural Network cont'd

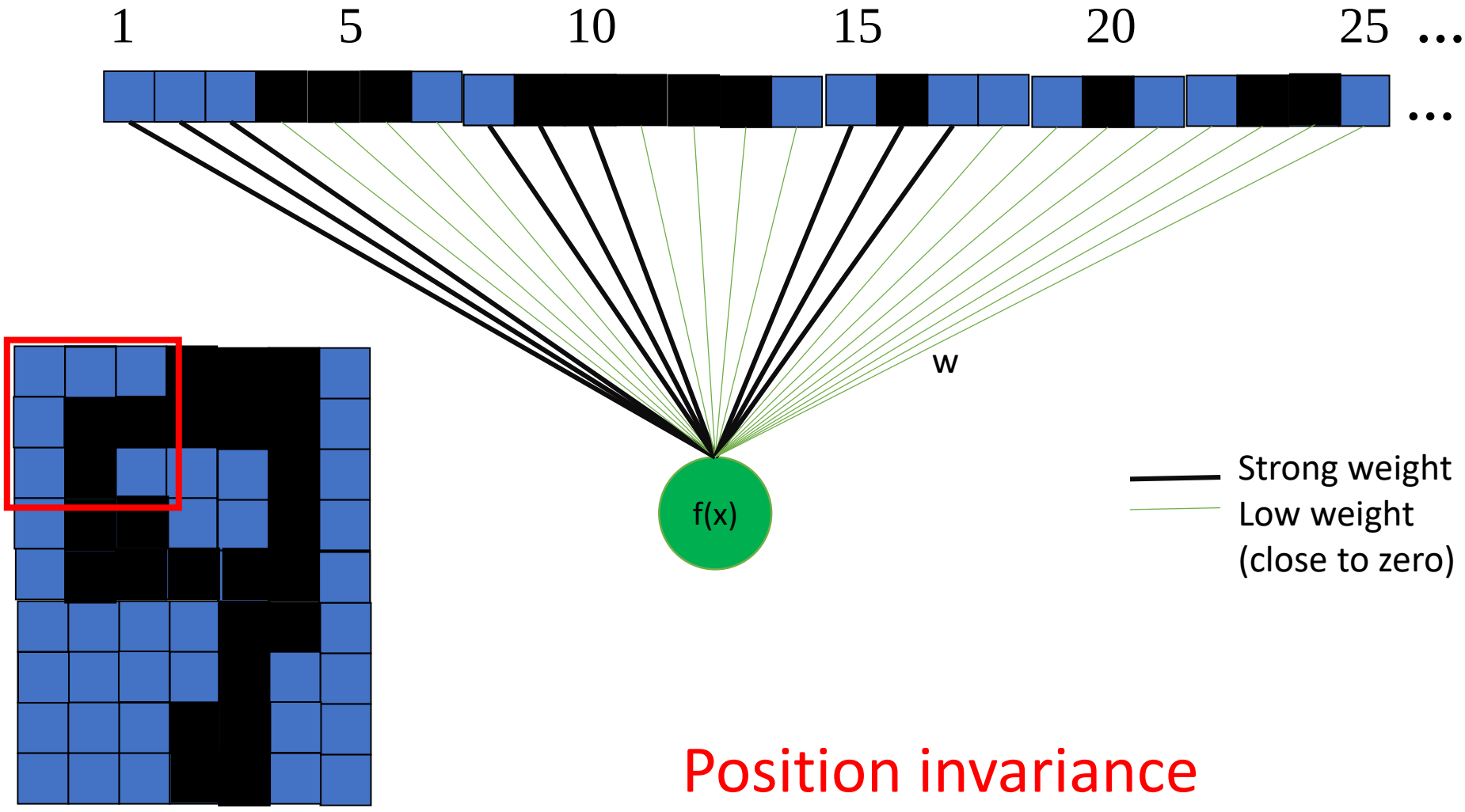


Neural Network cont'd



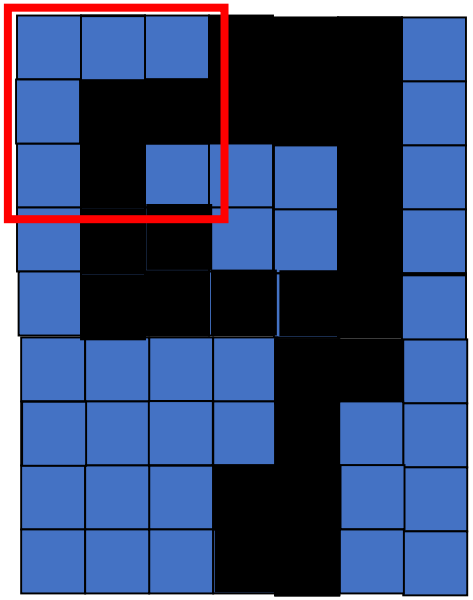


Neural Network cont'd

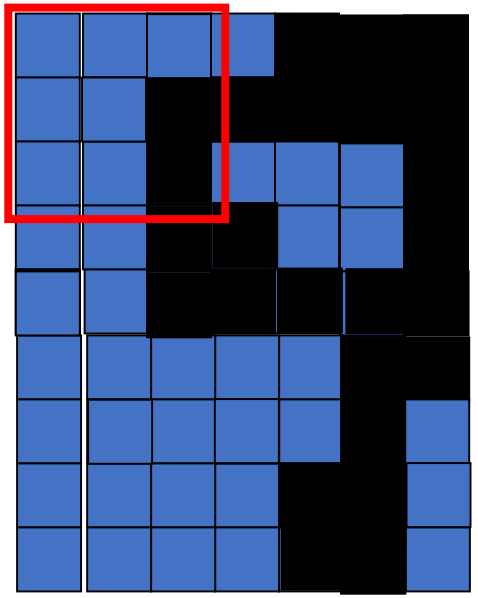


Position invariance

Neural Network cont'd



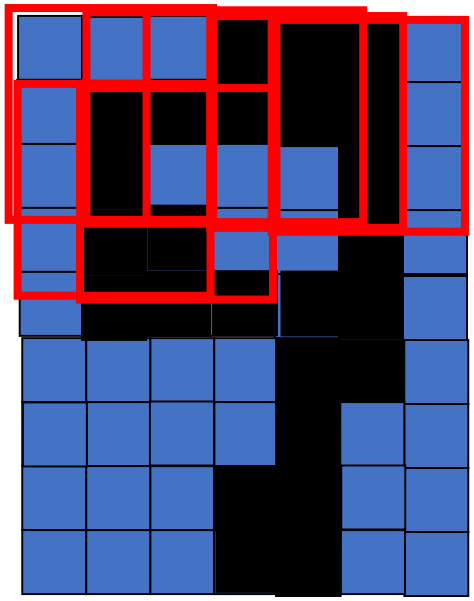
A training sample.



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

Position invariance

Convolutional Neural Network

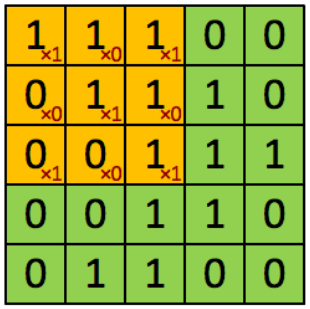


A training sample.

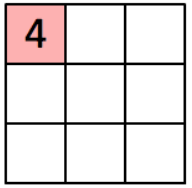
https://www.youtube.com/watch?v=f0t-OCG79-U

Convolutional Neural Network

- kernel = filter = feature detector
  - Mapping window

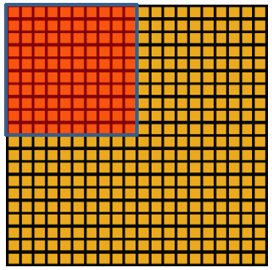


Image

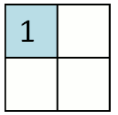


Convolved Feature

- Pooling
  - Reduce the window
  - Average, Max, etc.



Convolved feature

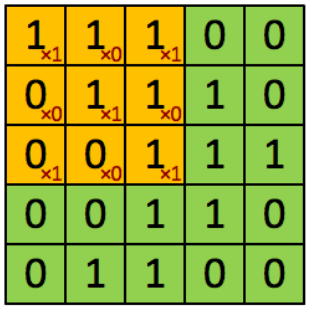


Pooled feature

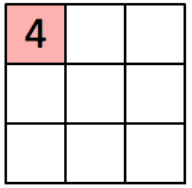
http://www.wildml.com/2015/11/understanding-convolutional-neural-networks-for-nlp/
http://deeplearning.stanford.edu/tutorial/supervised/Pooling/

Convolutional Neural Network

- kernel = filter = feature detector
  - Mapping window

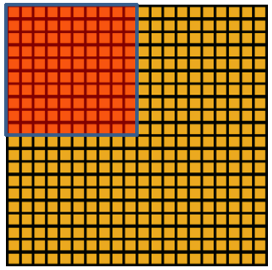


Image

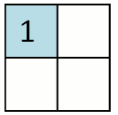


Convolved Feature

- Pooling
  - Reduce the window
  - Average, Max, etc.



Convolved feature



Pooled feature

http://www.wildml.com/2015/11/understanding-convolutional-neural-networks-for-nlp/
http://deeplearning.stanford.edu/tutorial/supervised/Pooling/

Convolutional Neural Network

- Features:
  - Position Invariance
  - Receptive field of different granularity