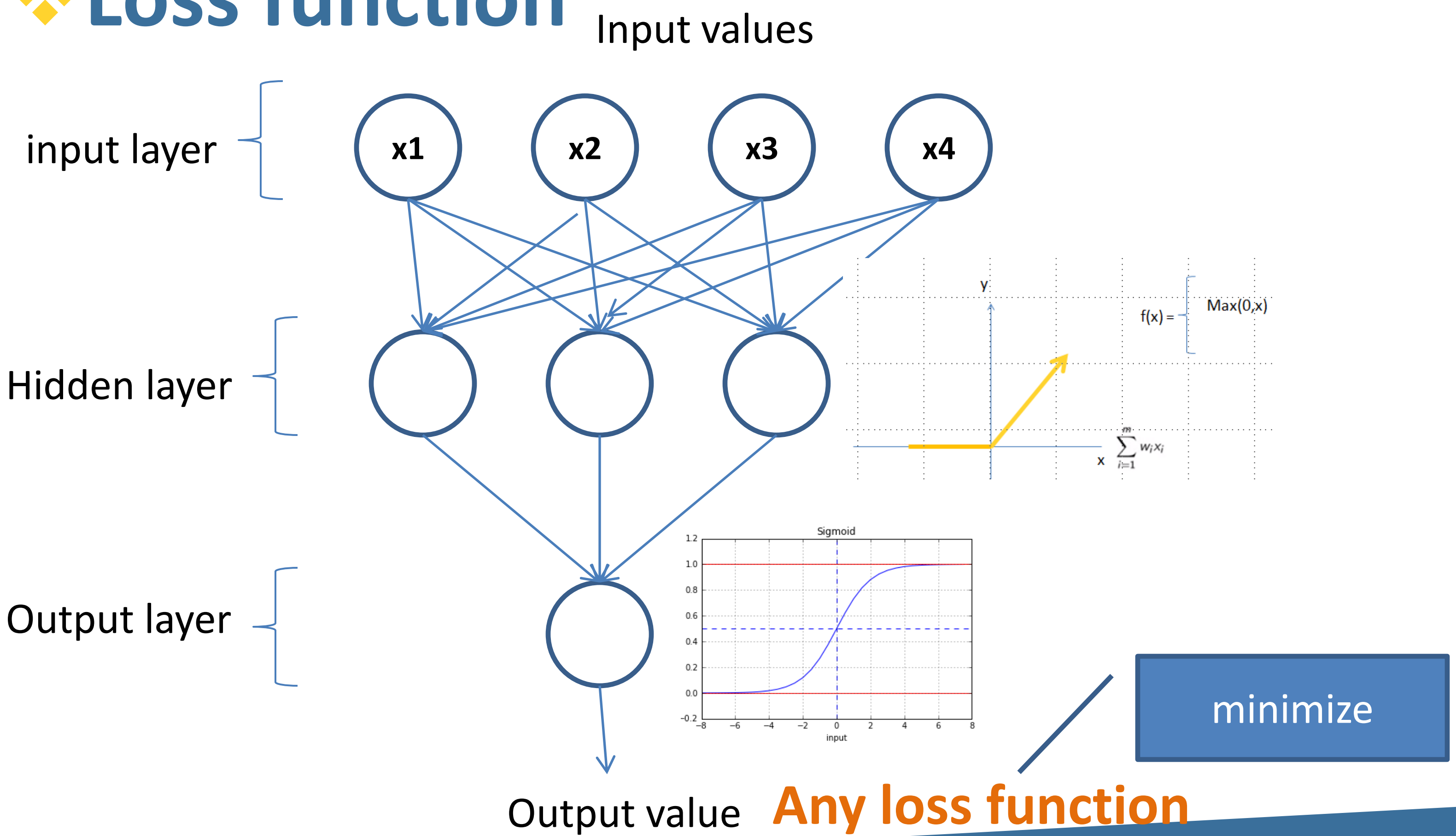
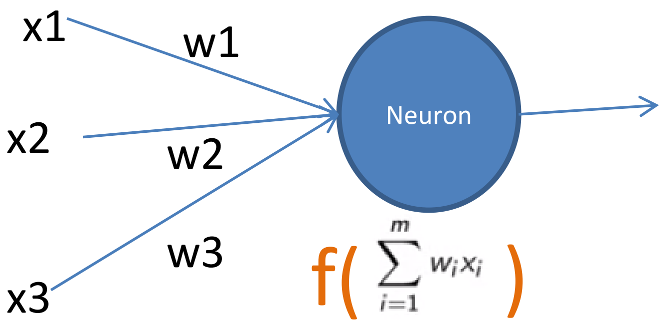
Loss function



Where is the learning?



Predicted value: y’

Actual value: y

Cost function: ½(y-y)^2

Steps

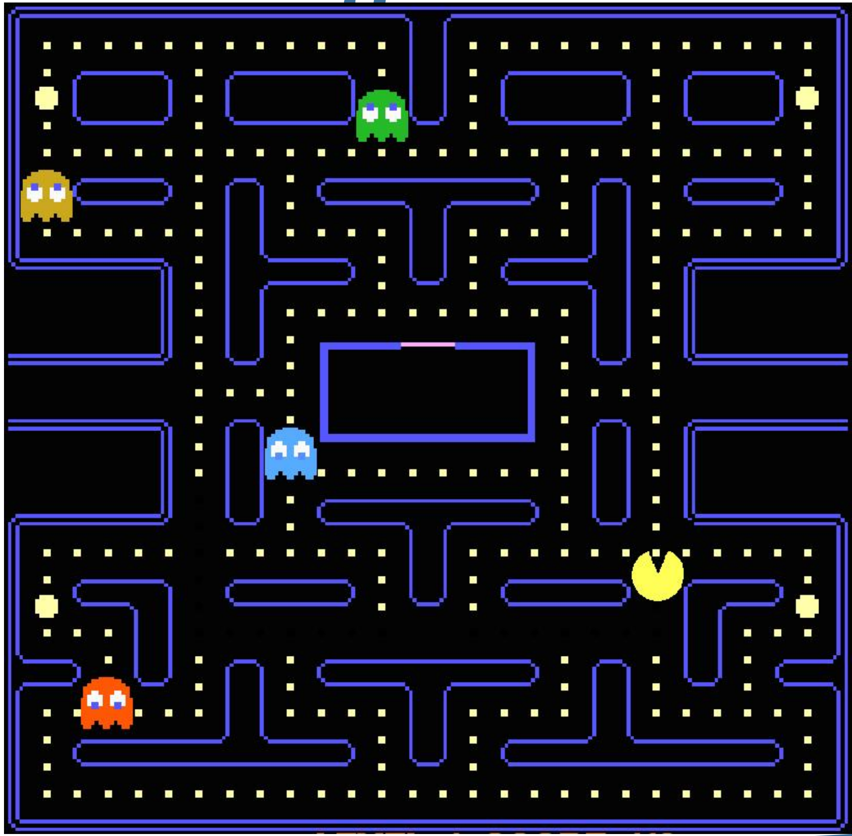
1. Random weights
2. Input observation to input layer. Each feature to one node
3. Forward propagation from left to right. Each neuron is activated based on the weights. Till you get an output
4. Compare the predicted result to the actual result. Measure the error (loss)
5. Back-propagate the error from right to left. Update the weights according to how much they are responsible
6. Repeat 1 – 5. Update the weights after each observation or each batch
7. Epoc is when the whole training set is passes through the ANN. Redo the epocs

Chapter: Reinforcement learning

Agenda

1) What is reinforcement learning
2) The bellman equation
3) Markov Decision Process
4) Policy vs Plan
5) Living Penalty
6) Q-Learning

What is reinforcement learning



Environment :

Any environment : for eg: Road, kitchen

Agent:

Object which works in the environment

Takes actions as performs steps

Gets a reward: Good reward for favorable state, bad reward for unfavorable state

What is reinforcement learning

Agent : Robot

Environment: Path

Actions that the agent can take:

stand,
lift right leg,
lift left leg,
plant right leg,
plant left get

Bellman Equation

Created by Richard Ernest Bellman

S – State

A- Action that the agent can take

R- Reward that the agent gets

Y – Discount

Bellman Equation

Bellman Equation

$$V(s) = \max(R(s,a) + \Upsilon V(s'))$$

S= current state

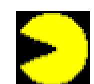
S’ = next state or where you end up

Bellman Equation

$$V(s) = \max(R(s,a) + \gamma V(s'))$$

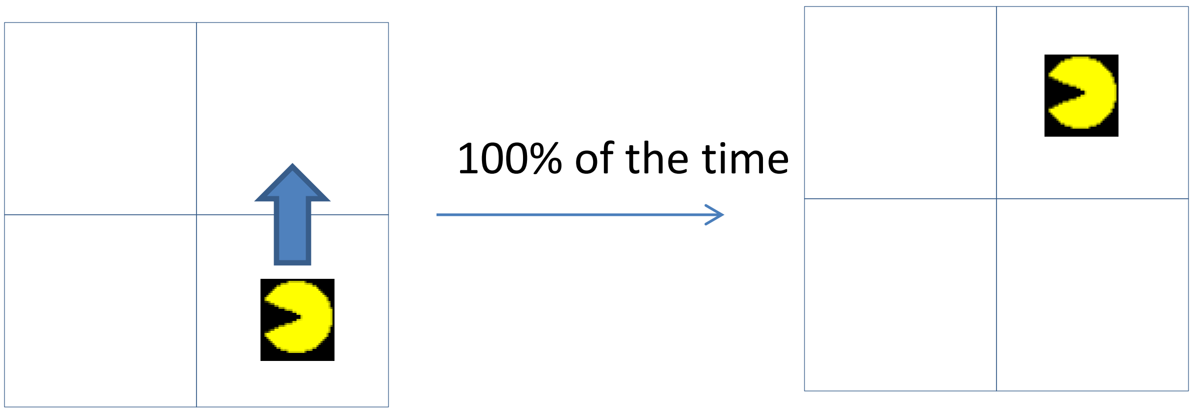
Bellman Equation – A Plan

	V=1	V=0.9	V=0.81
V=1	V=0.9	V=0.81	V=0.729
	V=0.81	V=0.729	V=0.6.6



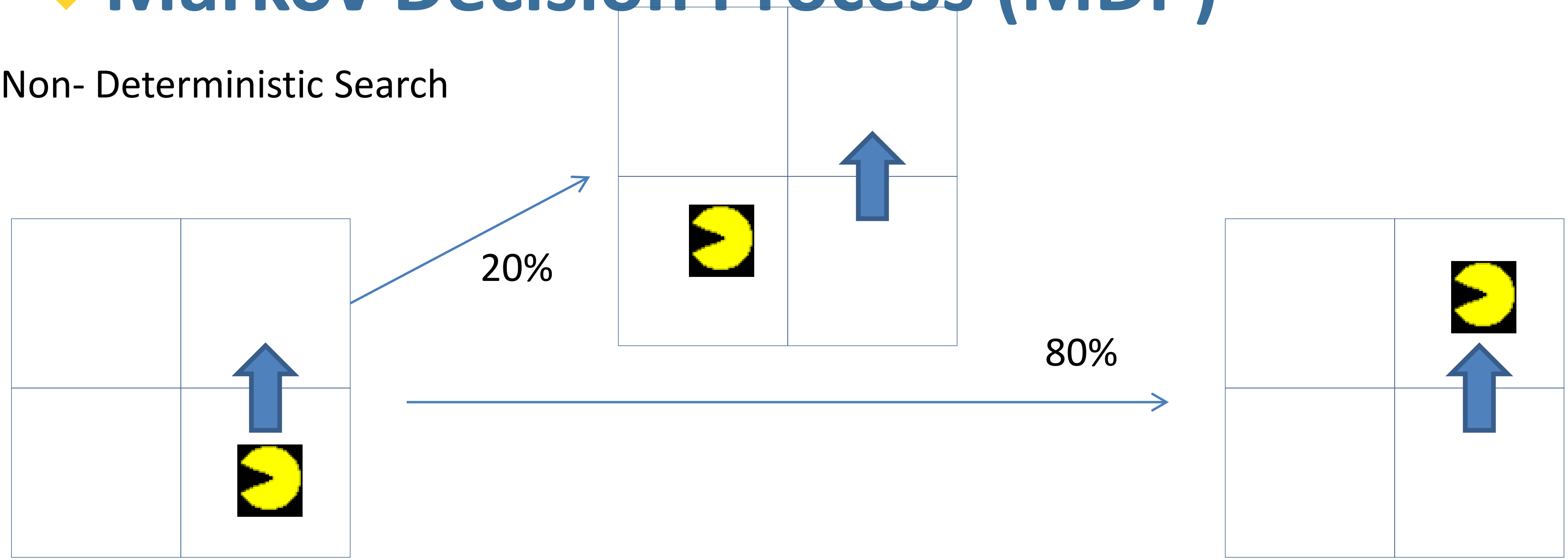
Markov Decision Process (MDP)

Deterministic Search



Markov Decision Process (MDP)

Non- Deterministic Search



Markov Process (MP)

A stocastic process has the Markov Property if the conditional probability distribution of the future states of the process depends only upon the present state , not on the sequence of events that preceded it. A process with this property is called a Markov process

Markov Decision Process (MP)

MDP provide a mathematical framework for modeling decision making in situations where outcomes are partly random and partly under the control of a decision maker

Bellman Equation

$$V(s) = \max(R(s,a) + \overbrace{YV(s')}^{0.8\,V(s1') \;+\; 0.2V(s2')})$$

$$V(s) = \max_a\left(R(s,a) + \gamma \sum_{s'} P(s,a,s')V(s')\right)$$

Policy Vs Plan

MDP provide a mathematical framework for modeling decision making in situations where outcomes are partly random and partly under the control of a decision maker

Chapter: Use Cases

Use cases

Smarter healthcare
Traffic control
JIT manufacturing
Trading analytics
Search quality
Customer behaviour
Fraud detection
Customer experience management
Predictive maintenance
Churn analysis
Security anomaly detection

Use Cases

Reinforcement Learning
- Autonomous systems

Generative
- Image enhancing

Predictive Maintenance
- networks
- Systems
- Human body