Model-Based Machine Learning

Chris Bishop
Microsoft Research
Cambridge UK

Machine learning algorithms

generative adversarial network

Markov random field

K-means clustering

Radial basis functions

decision trees

logistic regression

kernel PCA

Kalman filter

random forest

deep networks

principal components

Hidden Markov model

convolutional networks

support vector machines

Gaussian mixture

linear regression

Independent component analysis

neural networks

Gaussian process

factor analysis

Boltzmann machines

The ‘No Free Lunch’ Theorem

Averaged over all possible data-generating distributions, every classification algorithm has the same error rate when classifying previously unobserved points.

Wolpert (1996)

There is no universal machine learning algorithm

The goal of machine learning is to find an algorithm that is well matched to the problem being solved

Model-based machine learning

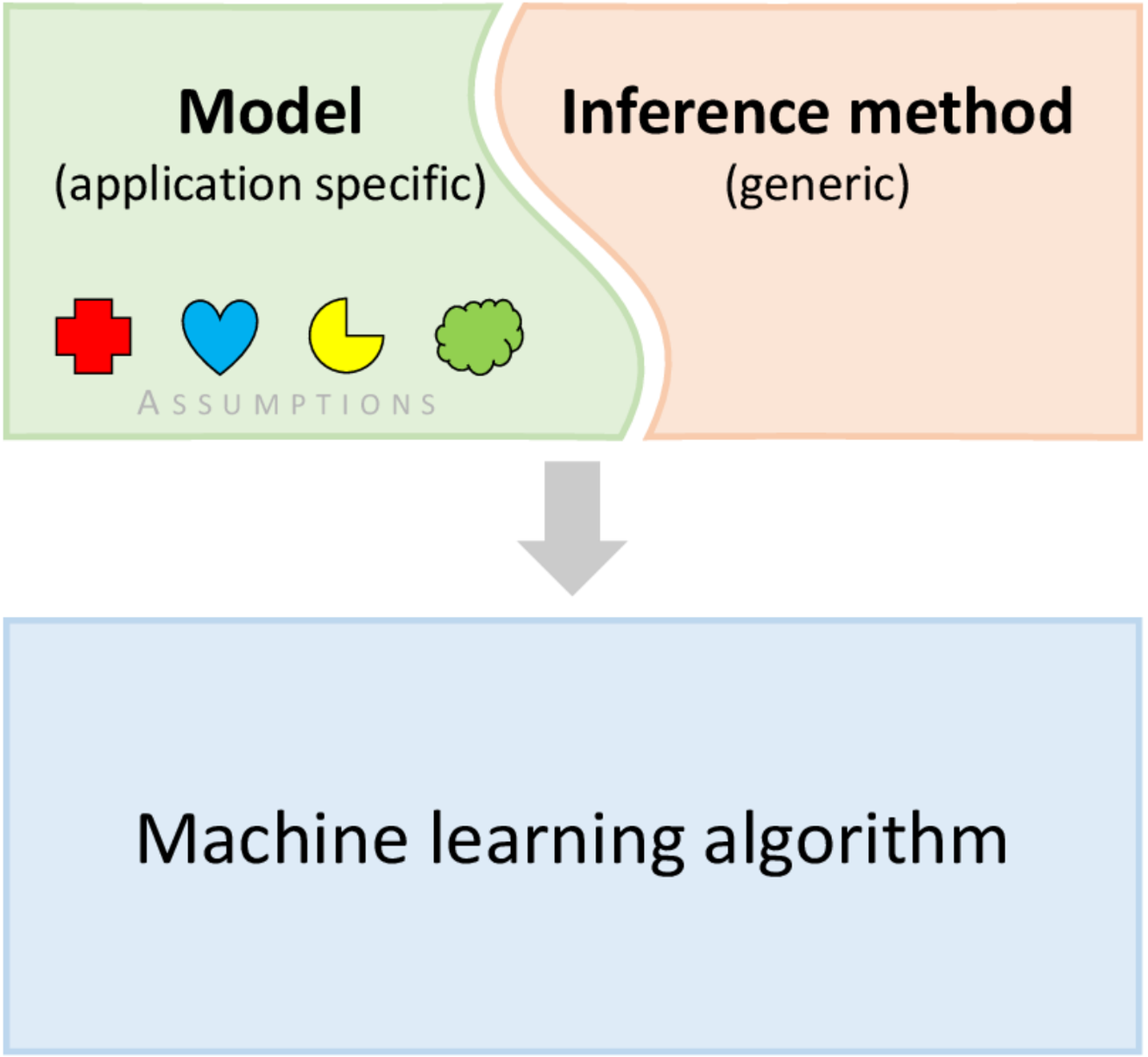
Derive the appropriate ML algorithm by making modelling assumptions explicit

Traditional:

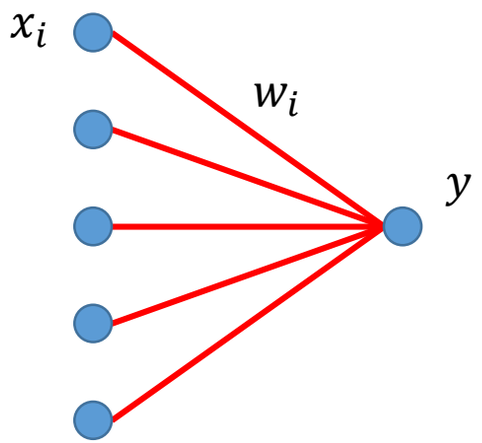
"how do I map my problem onto standard algorithms"?

Model-based:

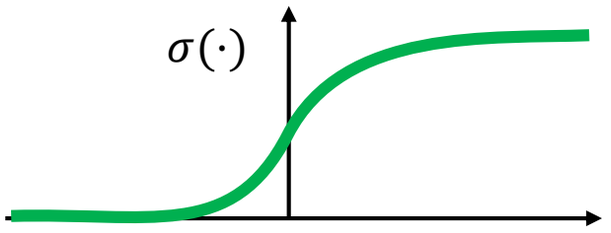
"what is the *model* that represents my problem"?


Model
(application specific)
ASSUMPTIONS
Inference method
(generic)
Machine learning algorithm

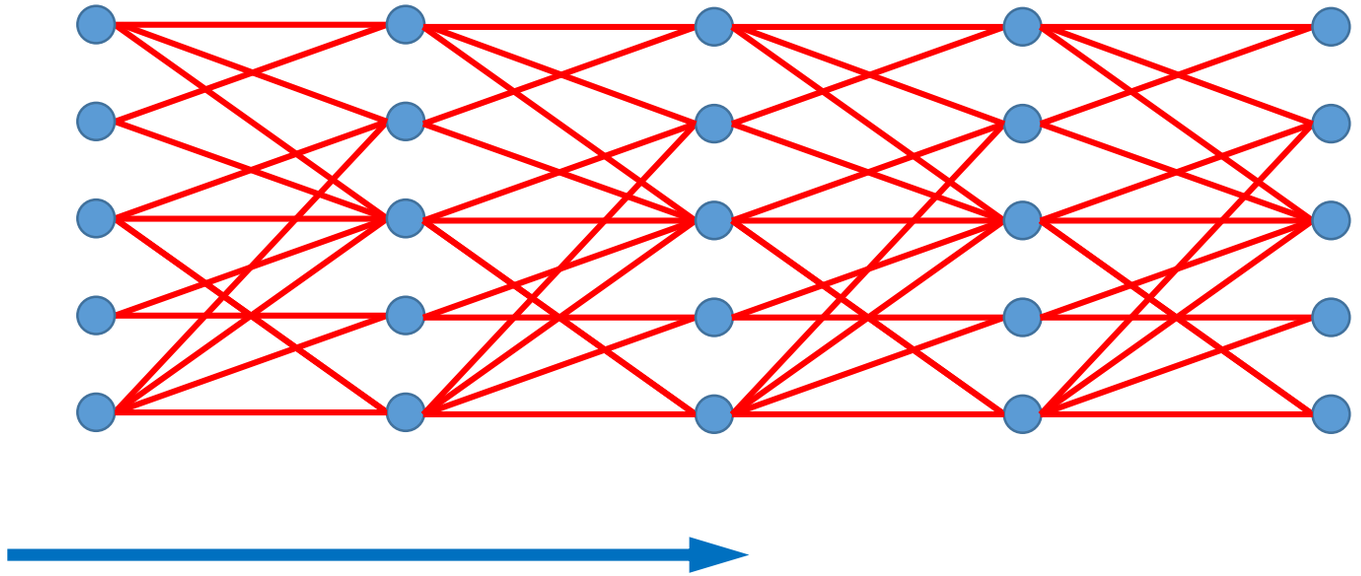
Logistic Regression



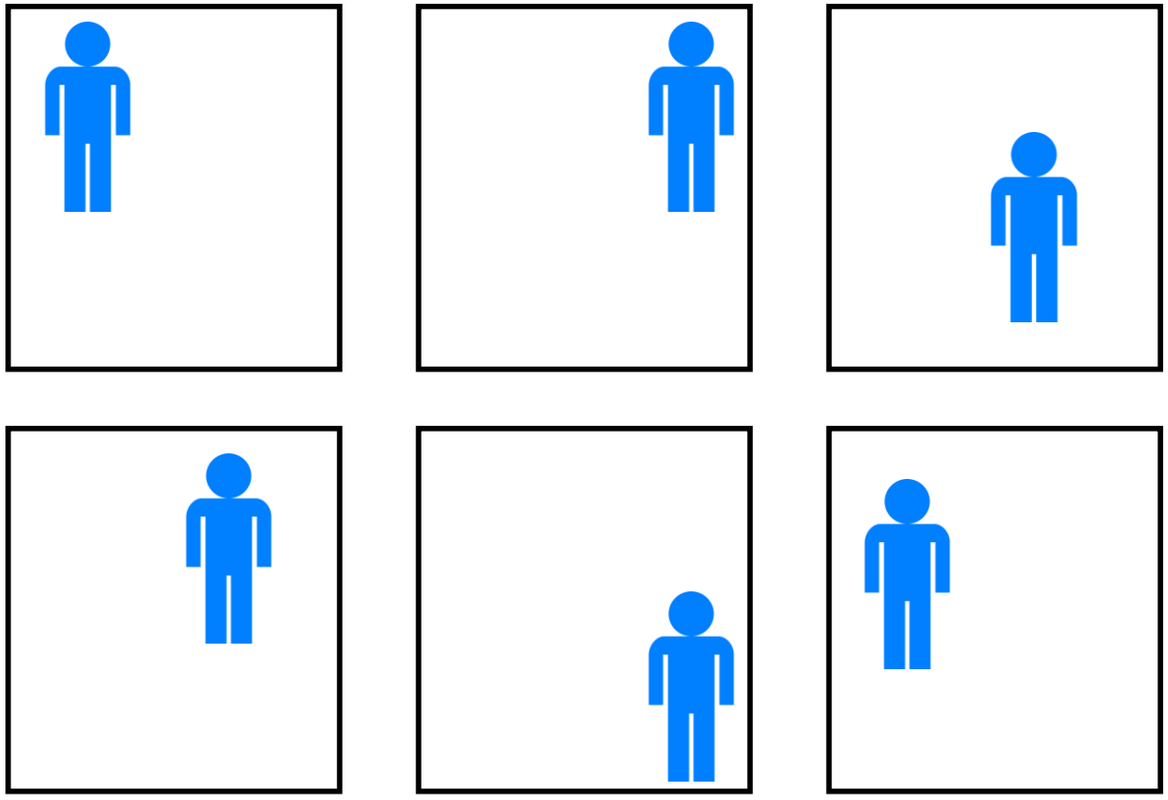
$$y = \sigma\left(\sum_i w_i x_i\right)$$

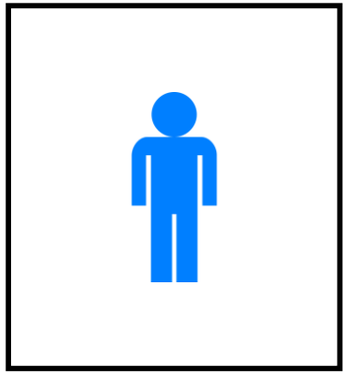


Deep Neural Networks



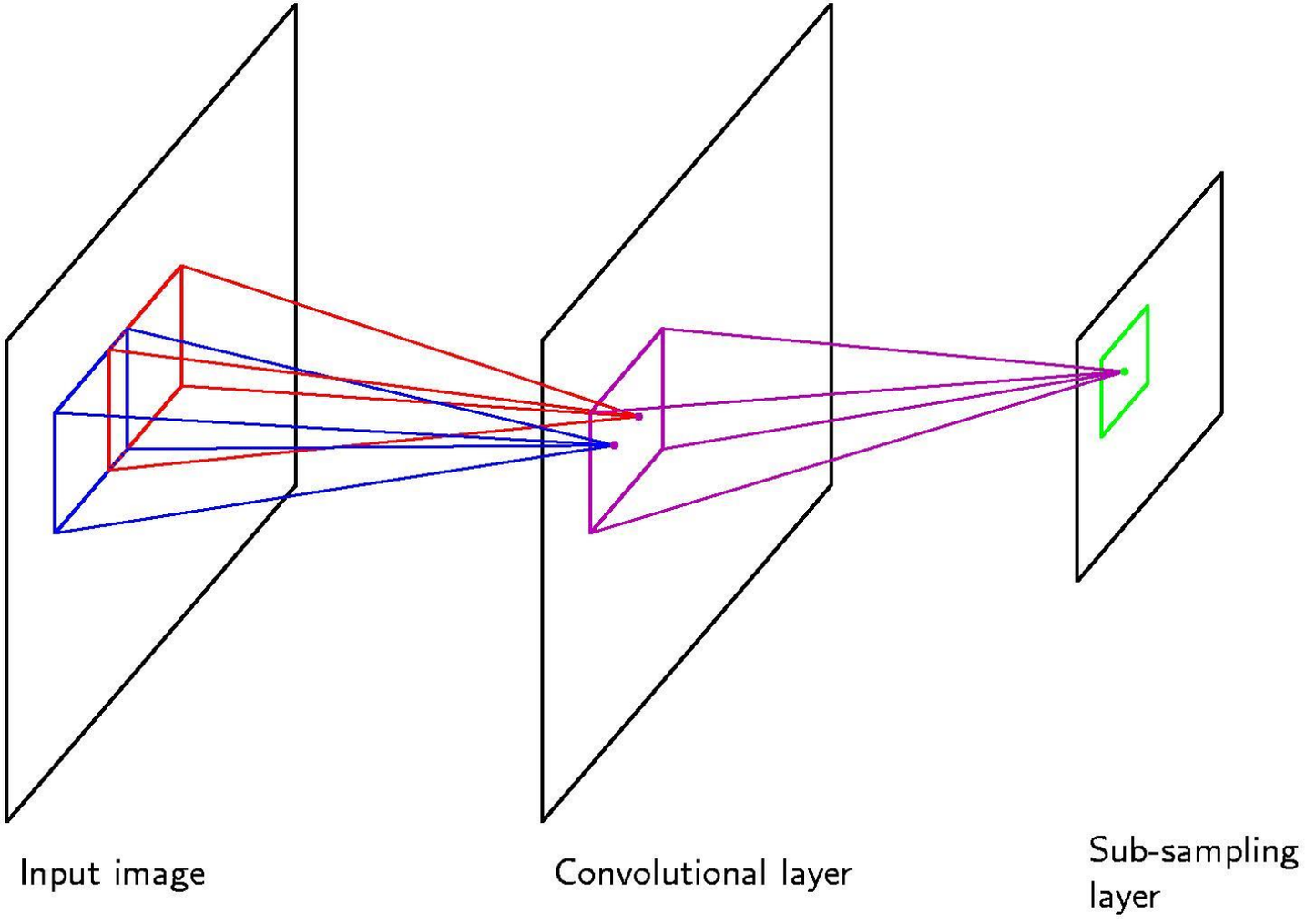
Data and prior knowledge



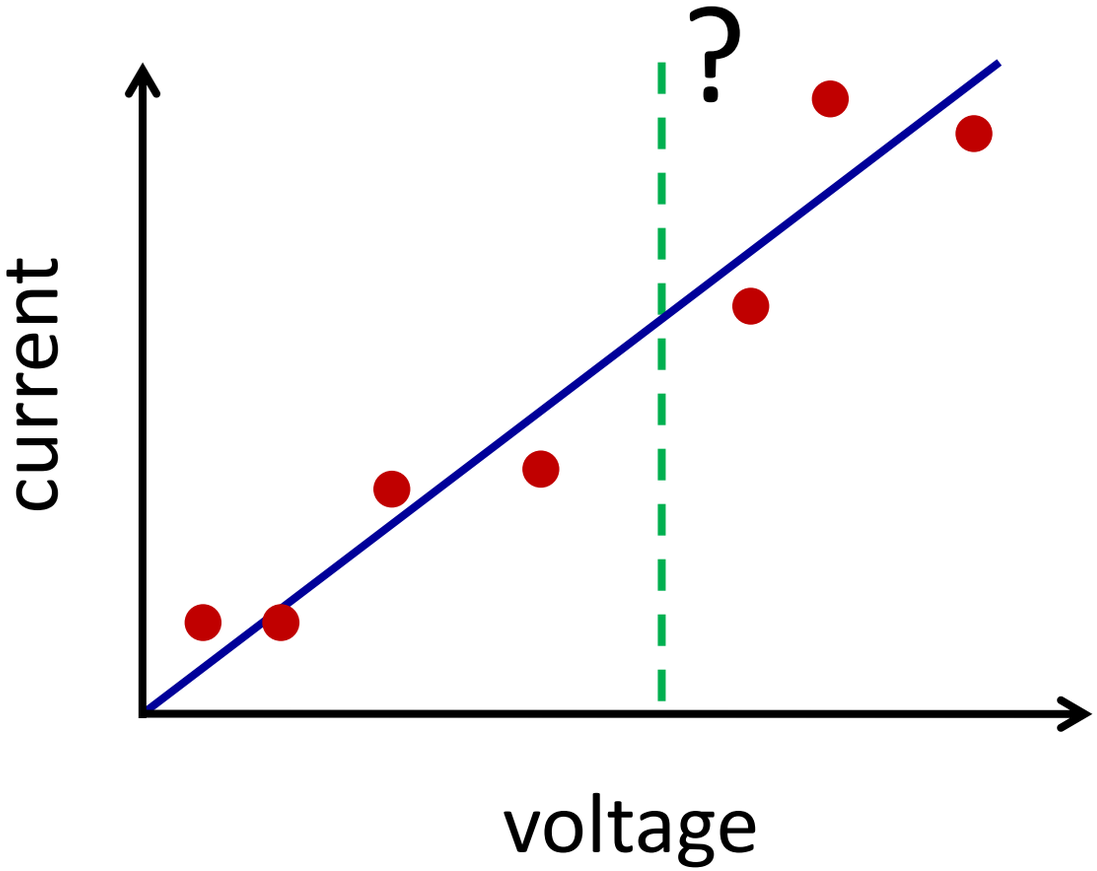


Translation invariance

Convolutional Neural Networks



‘Big data’



Uncertainty everywhere

Which movie should the user watch next?

Which word did the user write?

What did the user say?

Which web page is the user trying to find?

Which link will user click on?

Which gesture is the user making?

What is the prognosis for this patient?

Many others …

Probability

Limit of infinite number of trials (frequentist)

Quantification of uncertainty (Bayesian)

60%

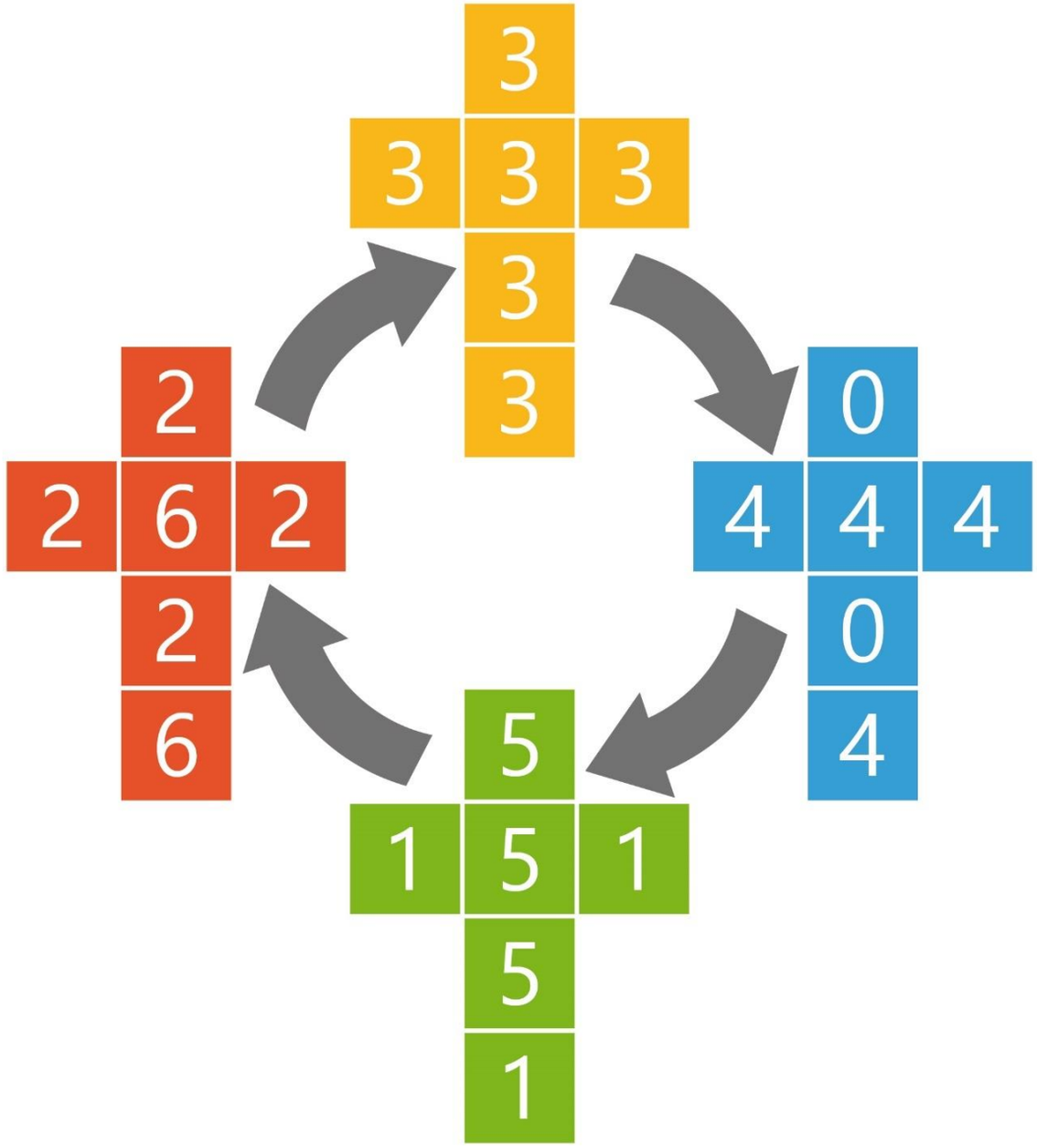
40%

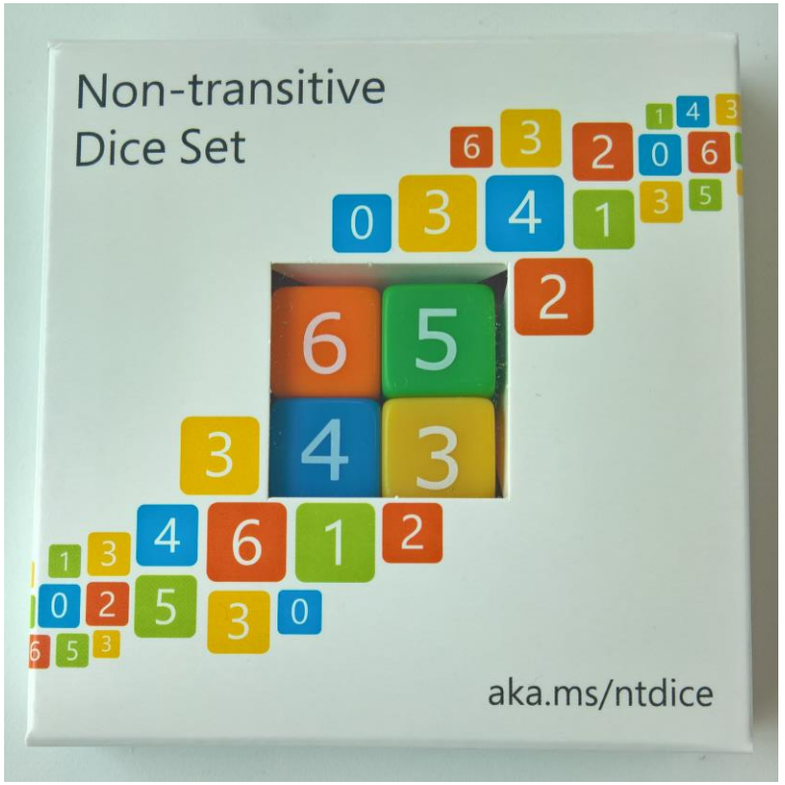

54
தமிழ்நாடு சொகுசுப் பேருந்து
BS III
PMI 294
ASHOK LEYLAND
TN-01-N-4347
PMI 294
மாநகர்பேருந்து


Fiesta


Lotsen

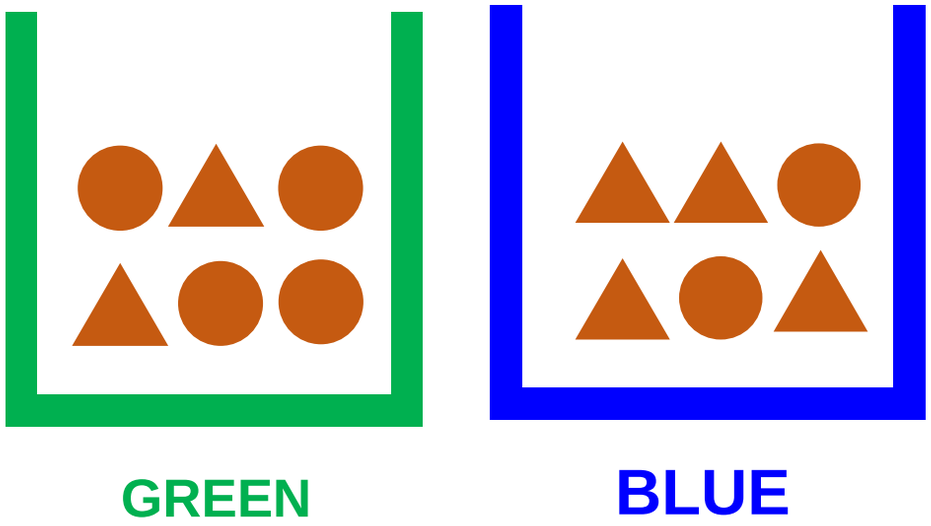

3
3
3
3
2
2
2
4
0
4
5
1
1

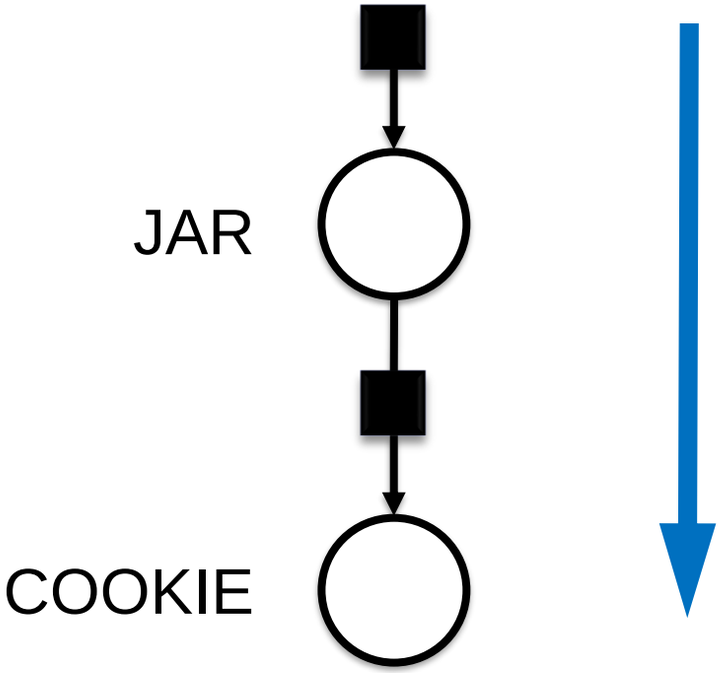


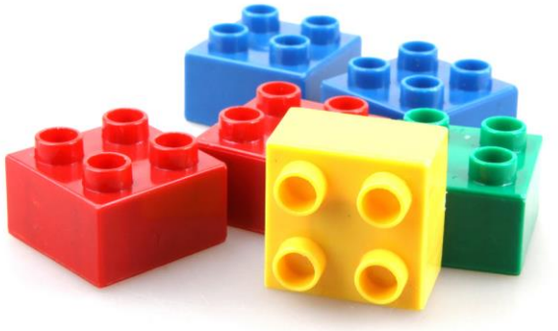


aka.ms/ntdice

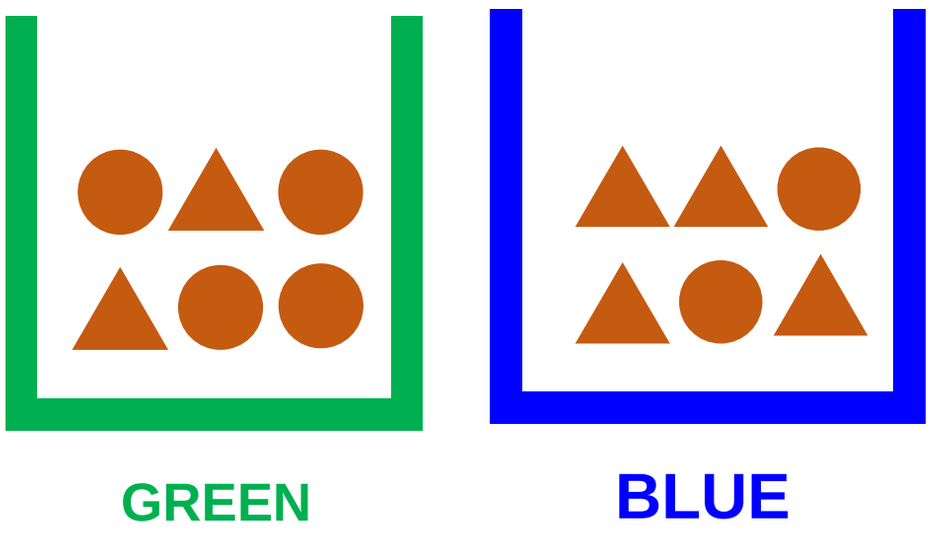
A Crash Course on Factor Graphs

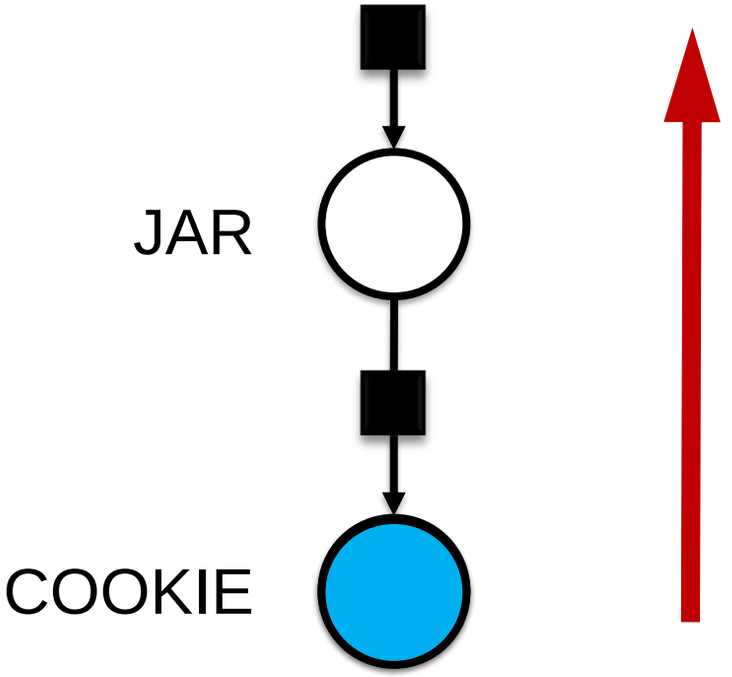


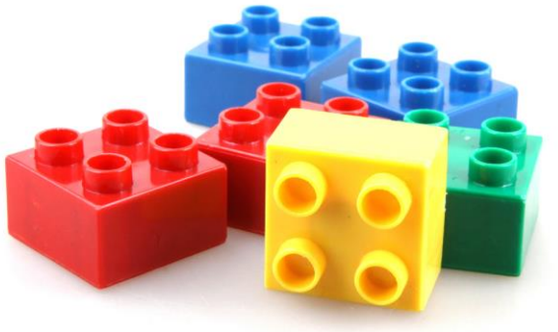




A Crash Course on Factor Graphs







PCA as an algorithm

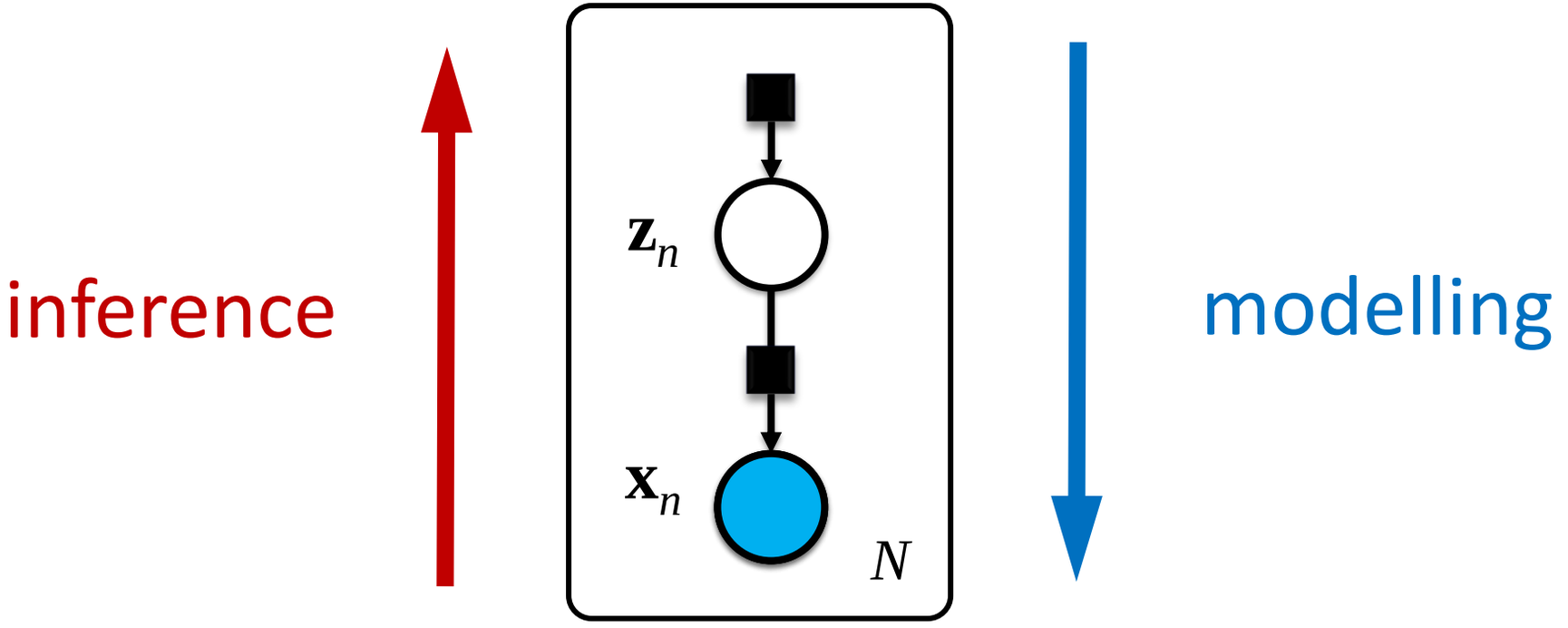
$$\overline{\mathbf{x}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}_n$$

$$\mathbf{S} = \frac{1}{N}\sum_{n=1}^{N}(\mathbf{x}_n - \overline{\mathbf{x}})(\mathbf{x}_n - \overline{\mathbf{x}})^{\mathrm{T}}$$

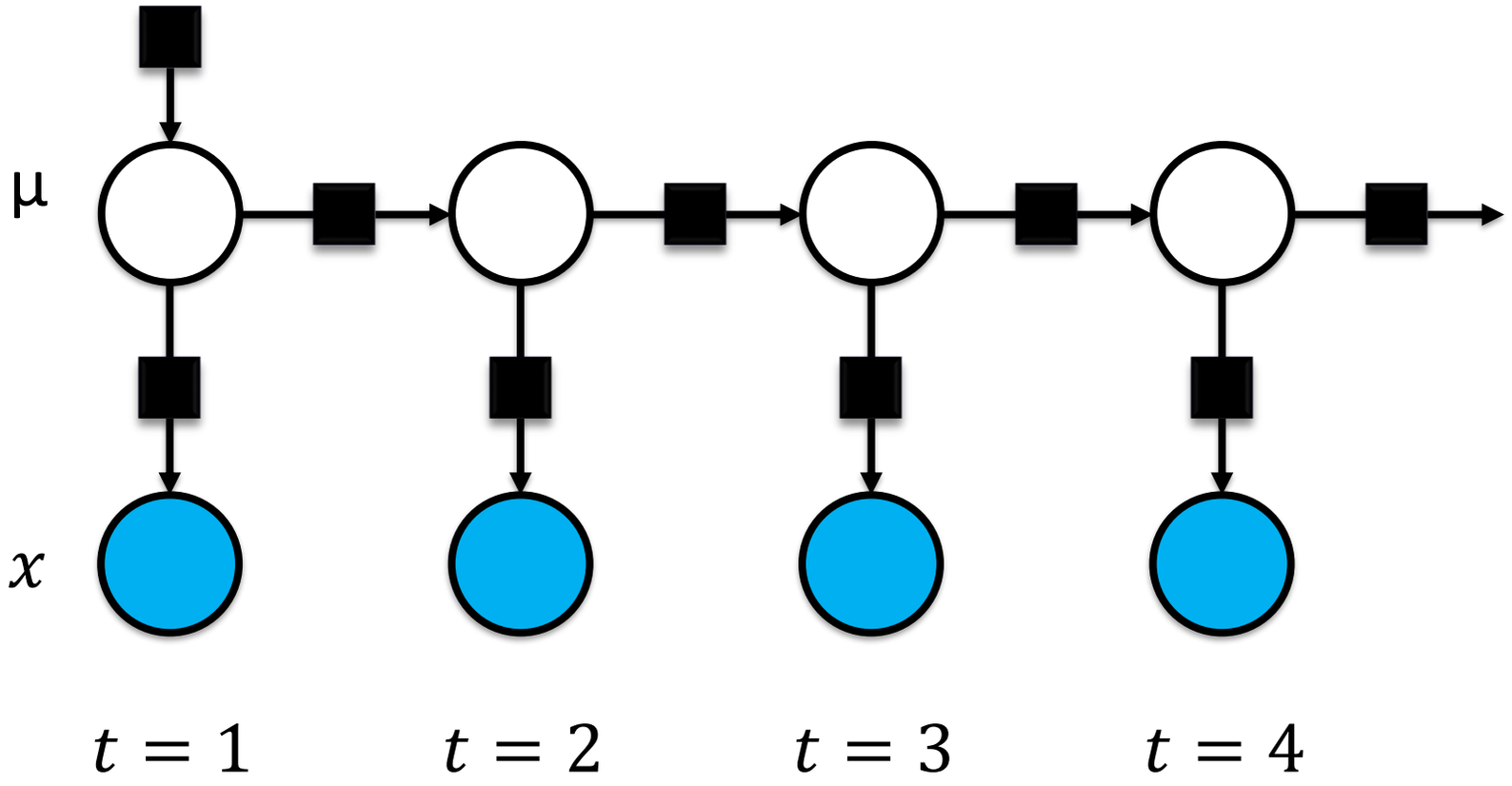
$$\mathbf{S}\mathbf{u}_i = \lambda_i \mathbf{u}_i$$

retain $M < D$ eigenvectors

PCA as a model



M. E. Tipping and C. M. Bishop (1997)

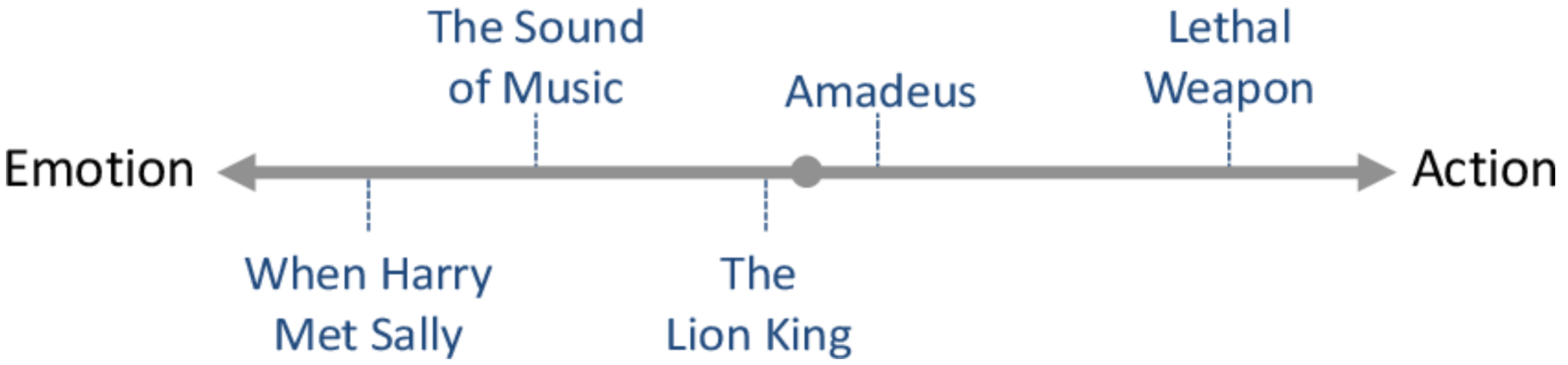


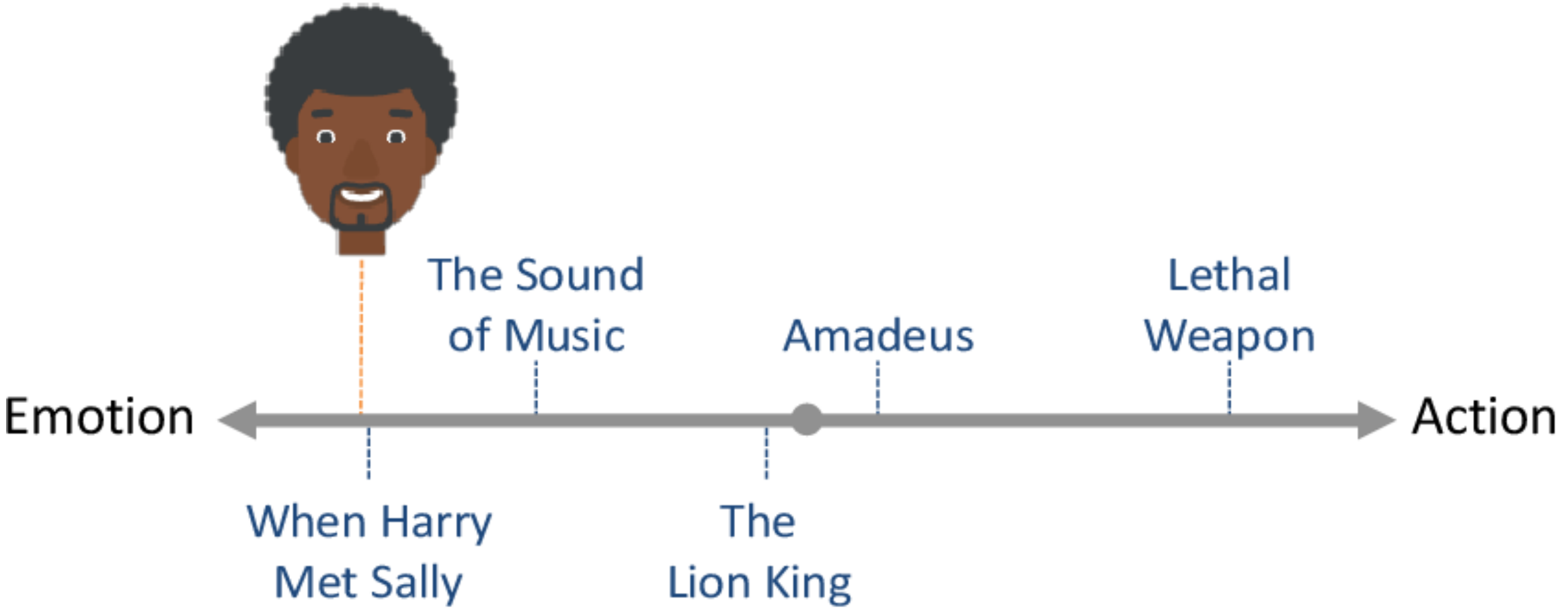
The Kalman filter

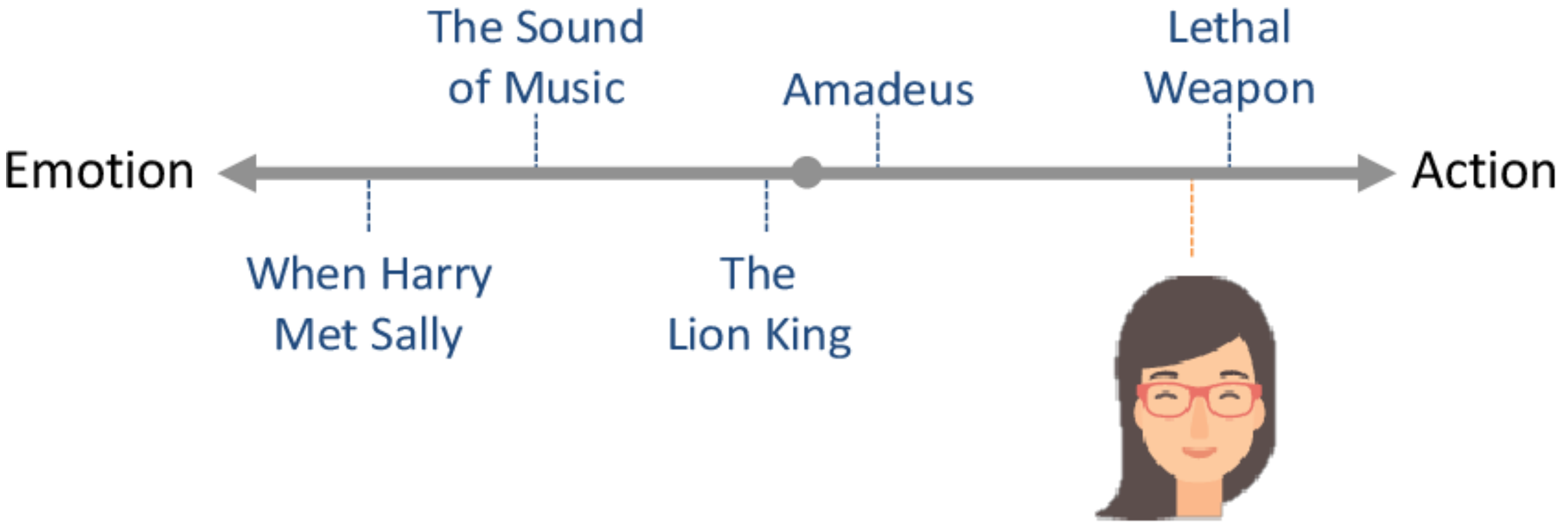
The hidden Markov model

Movie Recommender

Movie					
The Lion King	👍	?	?	👎	?
Lethal Weapon	👎	👍	👍	👍	?
The Sound of Music	?	👍	?	?	👎
Amadeus	👍	👎	?	👍	?
When Harry Met Sally	👍	👍	?	👎	👍

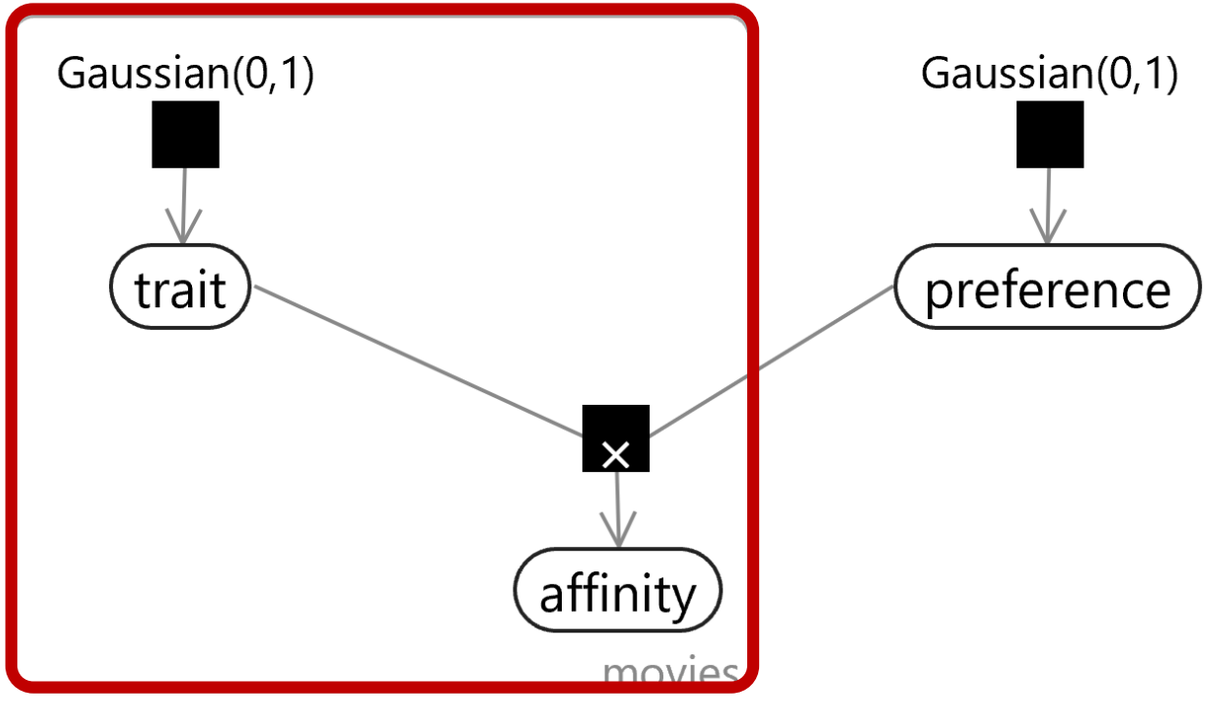

The Sound of Music
Amadeus
Lethal Weapon
Emotion
Action
When Harry Met Sally
The Lion King

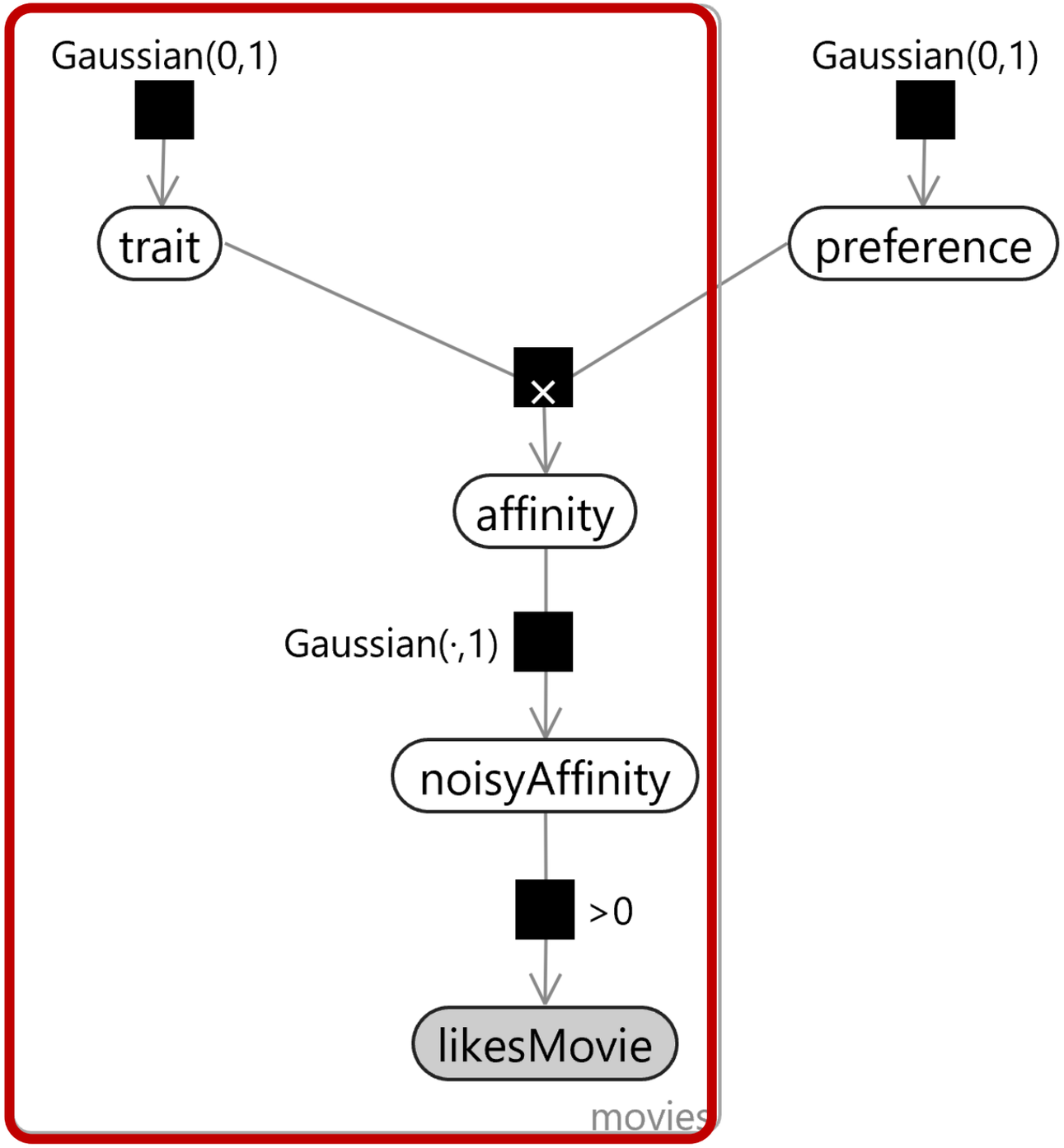

The Sound
of Music
Amadeus
Lethal
Weapon
Emotion
Action
When Harry
Met Sally
The
Lion King

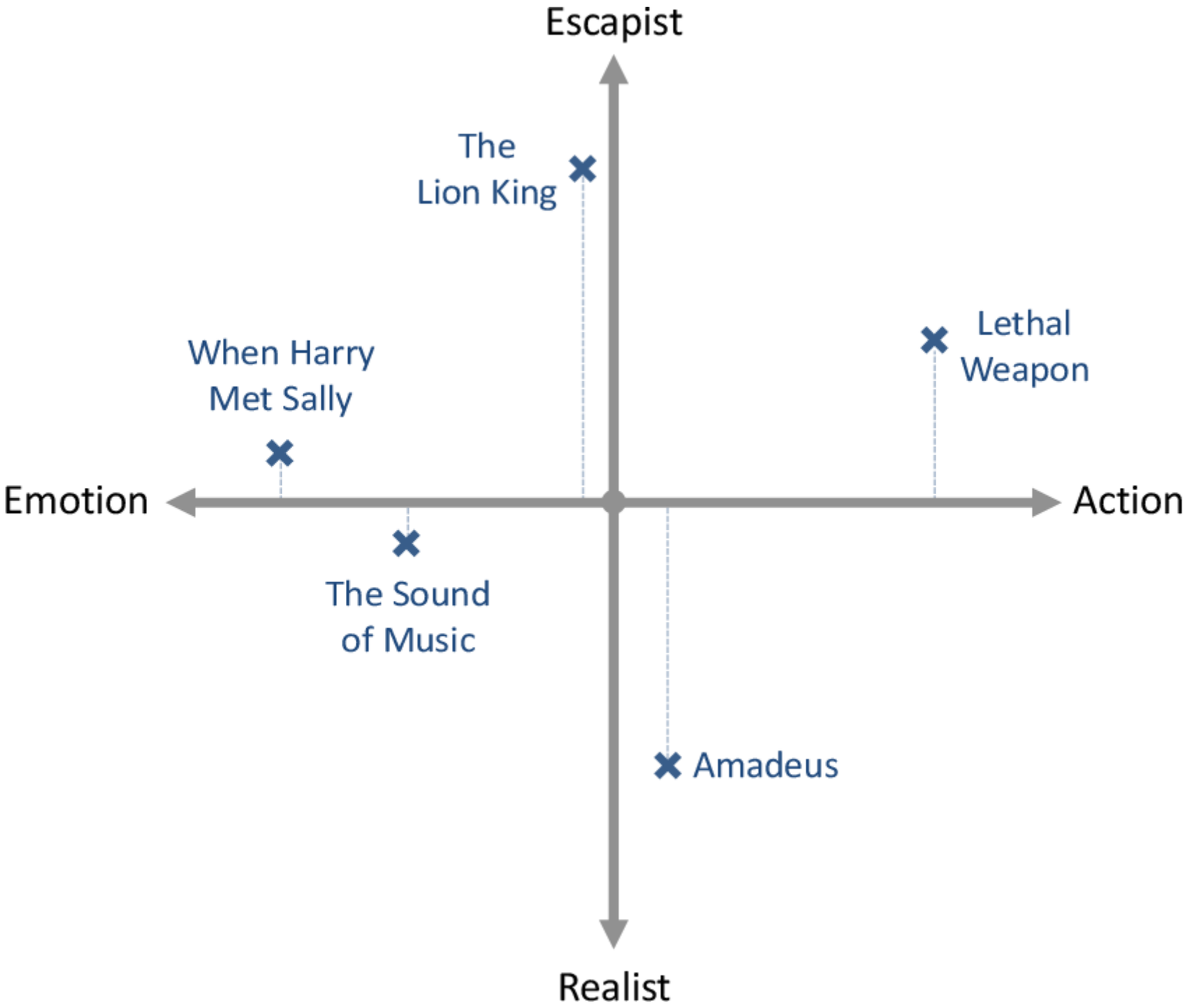

The Sound of Music
Amadeus
Lethal Weapon
Emotion
Action
When Harry Met Sally
The Lion King

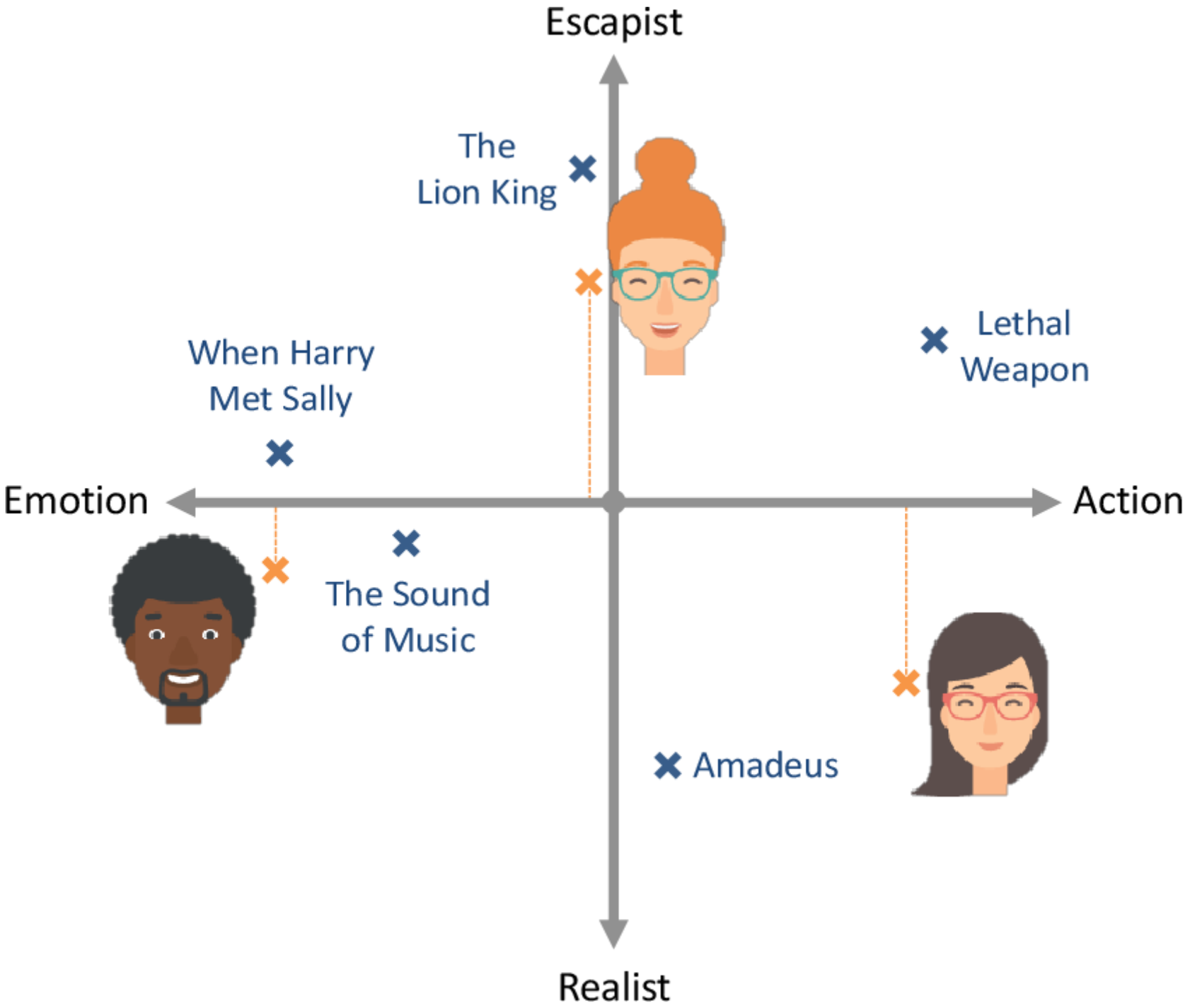
Assumptions

1) Each movie can be characterized by its position on the trait line, represented as a continuous number.
2) A person's preferences can be characterized by a position on the trait line, again represented as a continuous number.
3) A positive preference value means a person prefers movies with positive values of the trait (and vice versa for negative values). The absolute size of the preference value indicates the strength of preference.


Gaussian(0,1)
Gaussian(0,1)
trait
preference
affinity
movies

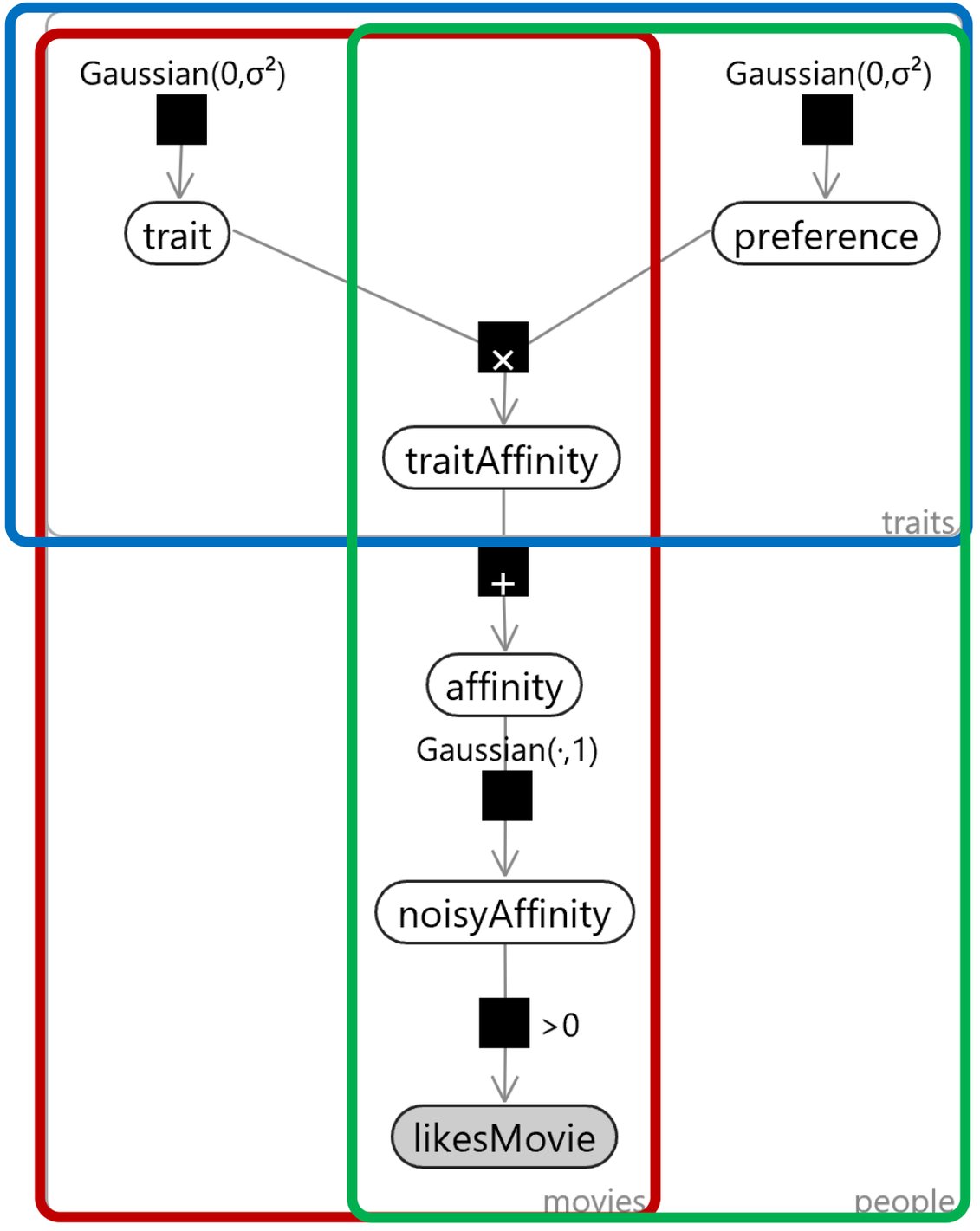

Gaussian(0,1)
Gaussian(0,1)
trait
preference
×
affinity
Gaussian(·,1)
noisyAffinity
>0
likesMovie
movies


Escapist
The
Lion King
When Harry
Met Sally
Lethal
Weapon
Emotion
Action
The Sound
of Music
Amadeus
Realist

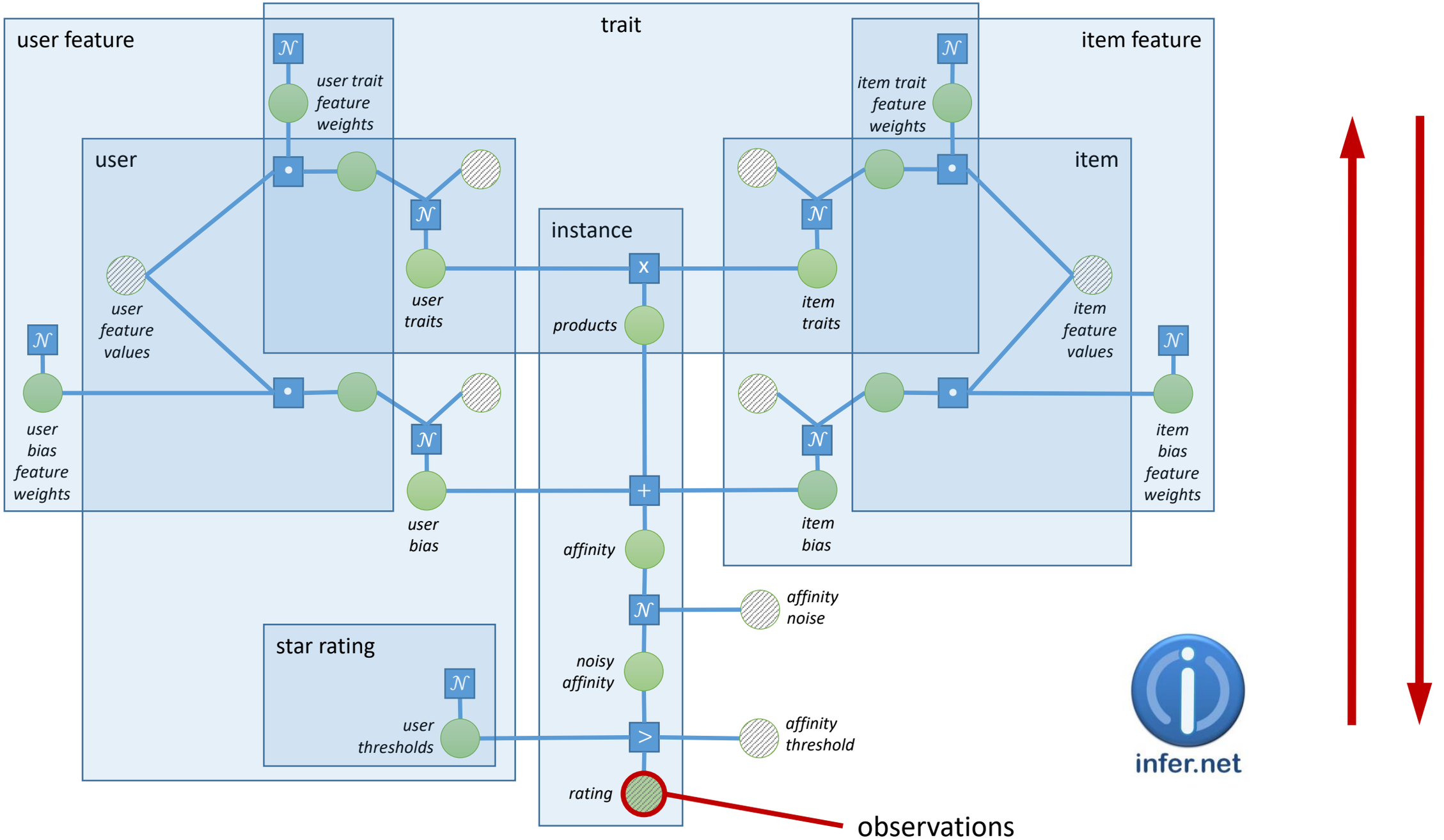

Escapist
The
Lion King
When Harry
Met Sally
Lethal
Weapon
Emotion
Action
The Sound
of Music
Amadeus
Realist

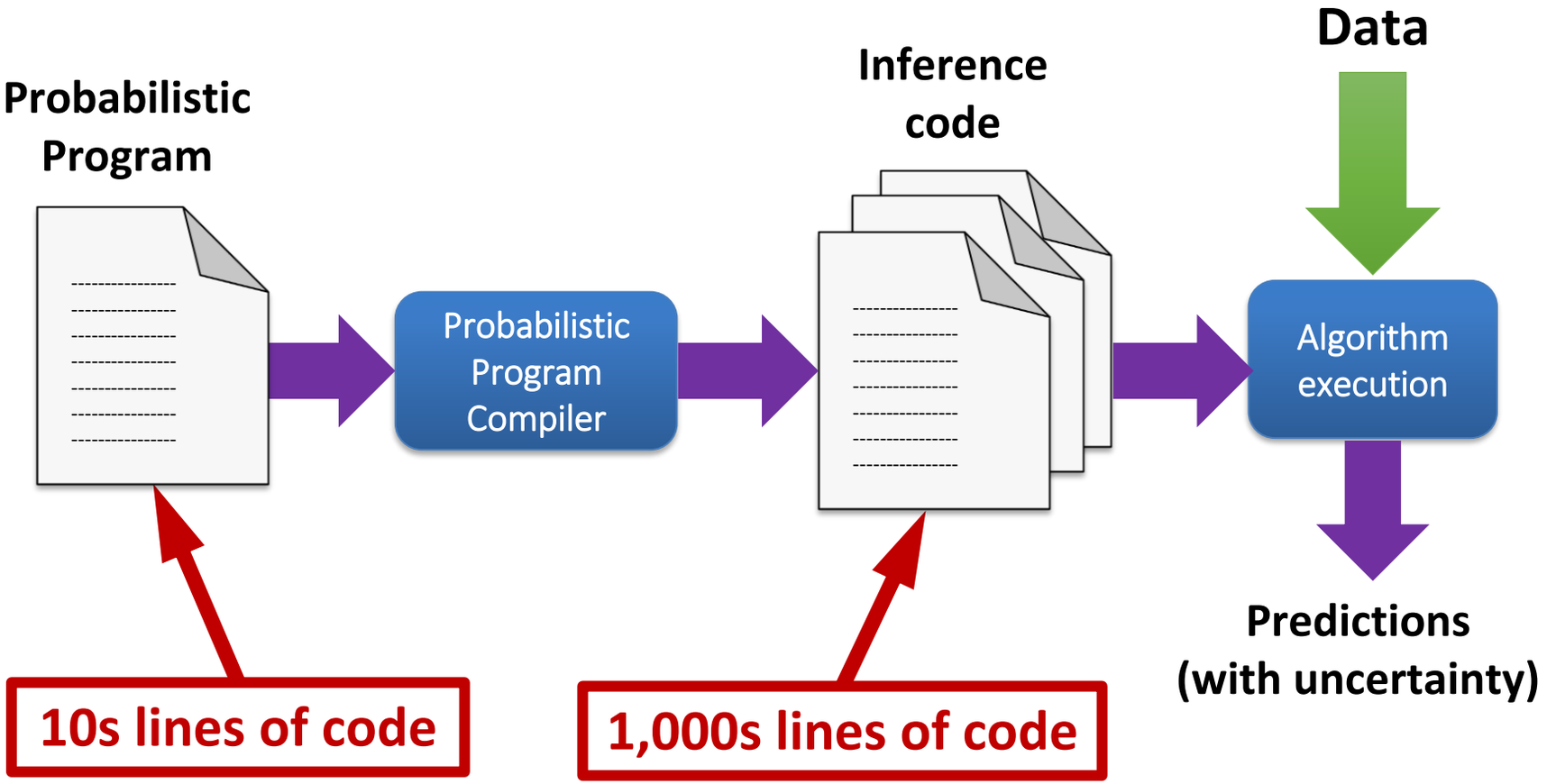
Assumptions

1) Each movie can be characterized by its position ~~on the trait line~~ in trait space, represented as a continuous number for each trait.
2) A person's preferences can be characterized by a position ~~on the trait line~~ in trait space, again represented as a continuous number for each trait.
3) A positive preference value means a person prefers movies with positive values of the trait (and vice versa for negative values). The absolute size of the preference value indicates the strength of preference.
4) The effect of one trait value on whether a person likes or dislikes a movie is the same, no matter what other trait values that movie has.
5) Whether a person will like or dislike a movie depends only on the movie's traits and not on anything else.


Gaussian(0,σ²)
Gaussian(0,σ²)
trait
preference
×
traitAffinity
traits
+
affinity
Gaussian(·,1)
noisyAffinity
>0
likesMovie
movies
people

Movie Recommender Demo


trait
user feature
item feature
user
item
instance
star rating
user trait feature weights
item trait feature weights
user feature values
user traits
item traits
item feature values
user bias feature weights
item bias feature weights
user bias
item bias
products
affinity
affinity noise
noisy affinity
affinity threshold
user thresholds
rating
observations
infer.net



http://infernet

infer.net

TEAM DEATHMATCH ON HARBOR

ROUND STARTING IN 0:05

RESPAWNS 15 ROUNDS TO WIN 2

	KILLS	DOWNS	REVIVES	DEATHS	SCORE	
I TwoSixNine I	--	--	--	--	--	0
III Agent T III	--	--	--	--	--	
jltarsenal	--	--	--	--	--	
Arekkz	--	--	--	--	--	
Oscar *	--	--	--	--	--	
Singoune	--	--	--	--	--	0
KEVLA89	--	--	--	--	--	
Feron Taylor	--	--	--	--	--	
Ix Fahrenheit x	--	--	--	--	--	
davkan	--	--	--	--	--	

ACTIVE BOUNTY

Rampage 0 / 1,000

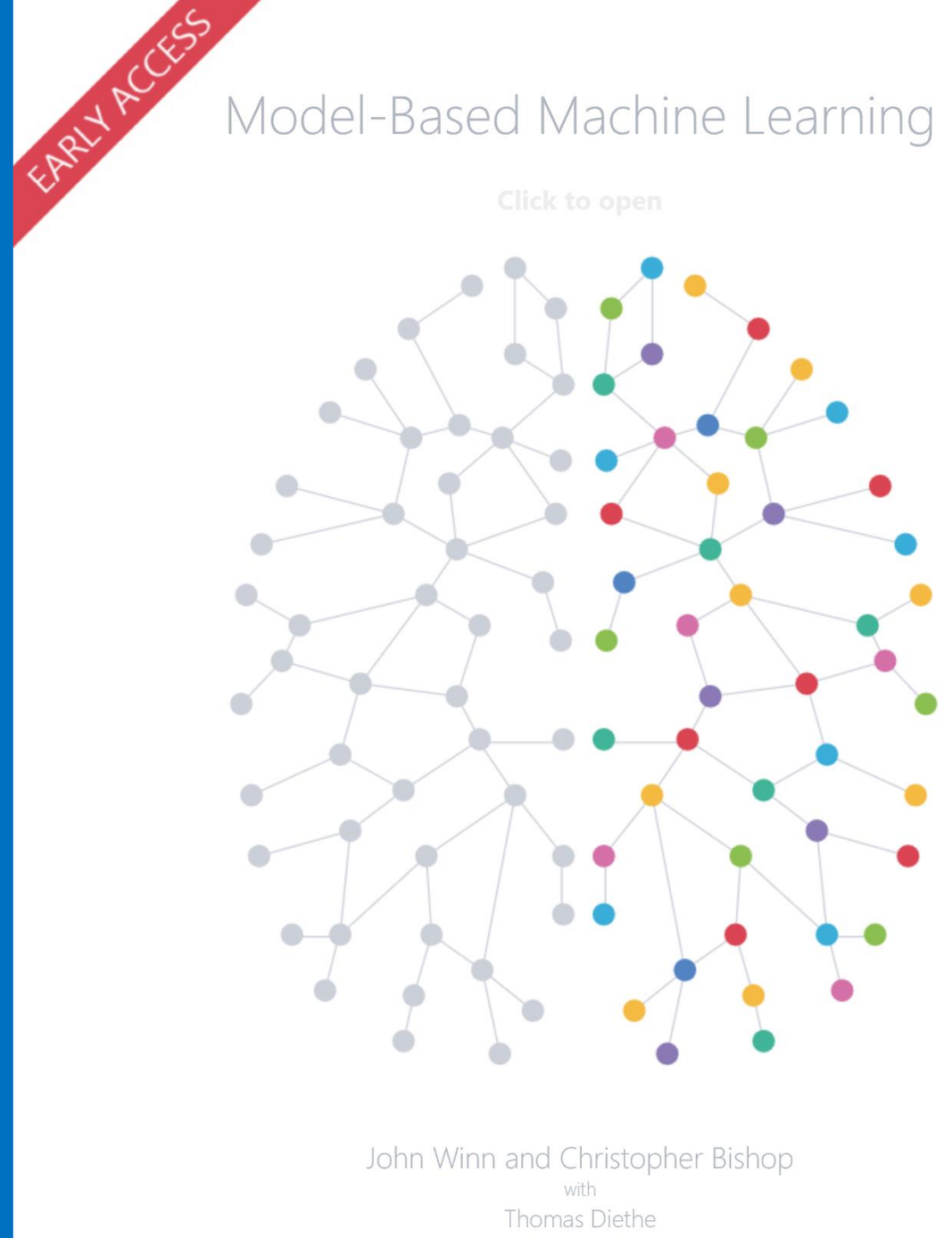
A MUTE PLAYER X VIEW GAMERCARD B OVERHEAD MAP

Y CHANGE LOADOUT

mbmlbook.com

Feedback please:

mbmlbook@microsoft.com



Thank you!