

# Statistical models

# Statistical modeling

Simplified mathematical formulation for statistical analysis

- stochastic
- somehow approximate reality
- explanatory

# Different philosophies

- Generative models
  - describe how the data are generated
- Predictive models
  - models mainly for making predictions

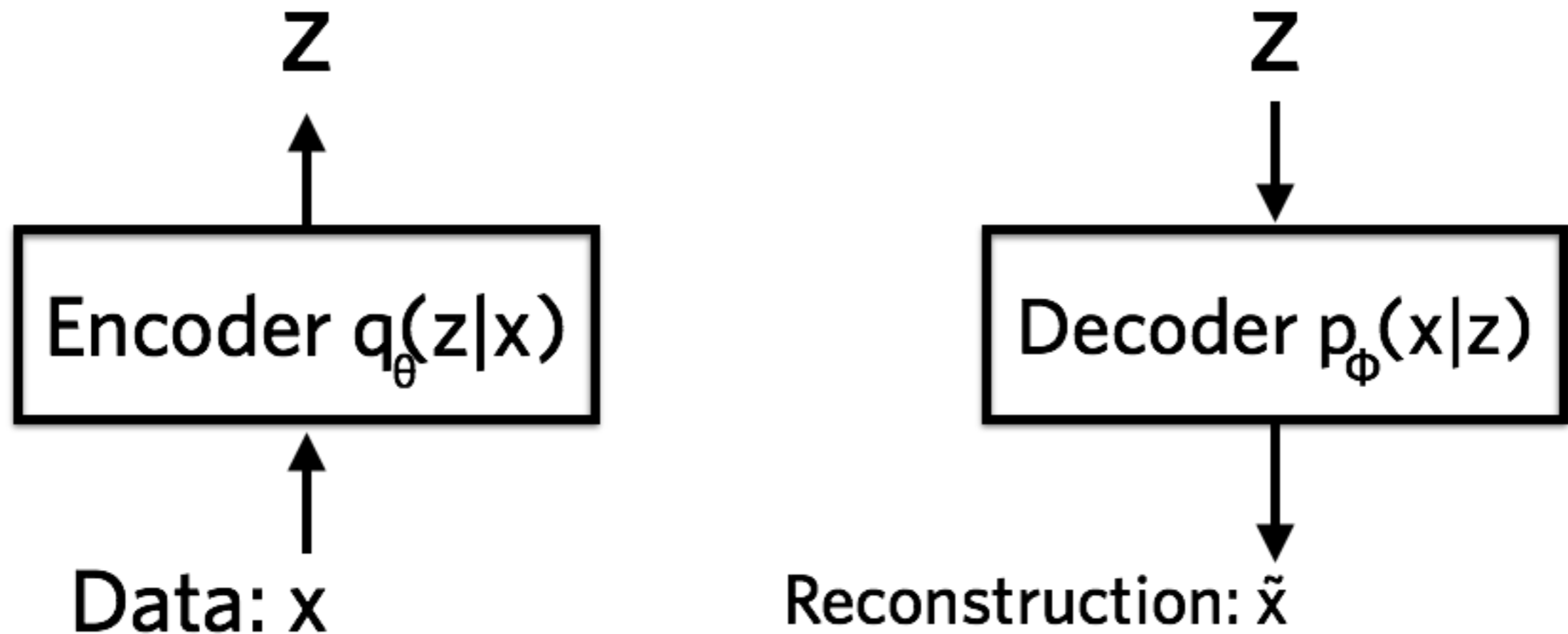
# Generative models

# Coin tossing

$X_i = \begin{cases} 1, & \text{with prob. } p \text{ (第 } i \text{ 次出現正面)} \\ 0, & \text{" } 1-p \text{ ( " 反面)} \end{cases}$

$\Rightarrow X_i \stackrel{\text{i.i.d.}}{\sim} \text{Bernoulli}(p)$

# Variational autoencoder





# Variational autoencoder





# Failure rate model



# State-space models for air pollution

$$P_{t+1}(S) = \int D_{\Theta}(S, S') P_t(S') dS' + N(S)$$

depend on weather

diffusion

新污染 or 沉降

# Predictive models

# Image recognition



**mite**

**container ship**

**motor scooter**

**leopard**

	<b>mite</b>		<b>container ship</b>		<b>motor scooter</b>		<b>leopard</b>
	<b>black widow</b>		<b>lifeboat</b>		<b>go-kart</b>		<b>jaguar</b>
	<b>cockroach</b>		<b>amphibian</b>		<b>moped</b>		<b>cheetah</b>
	<b>tick</b>		<b>fireboat</b>		<b>bumper car</b>		<b>snow leopard</b>
	<b>starfish</b>		<b>drilling platform</b>		<b>golfcart</b>		<b>Egyptian cat</b>

[https://www.tensorflow.org/tutorials/images/image\\_recognition](https://www.tensorflow.org/tutorials/images/image_recognition)

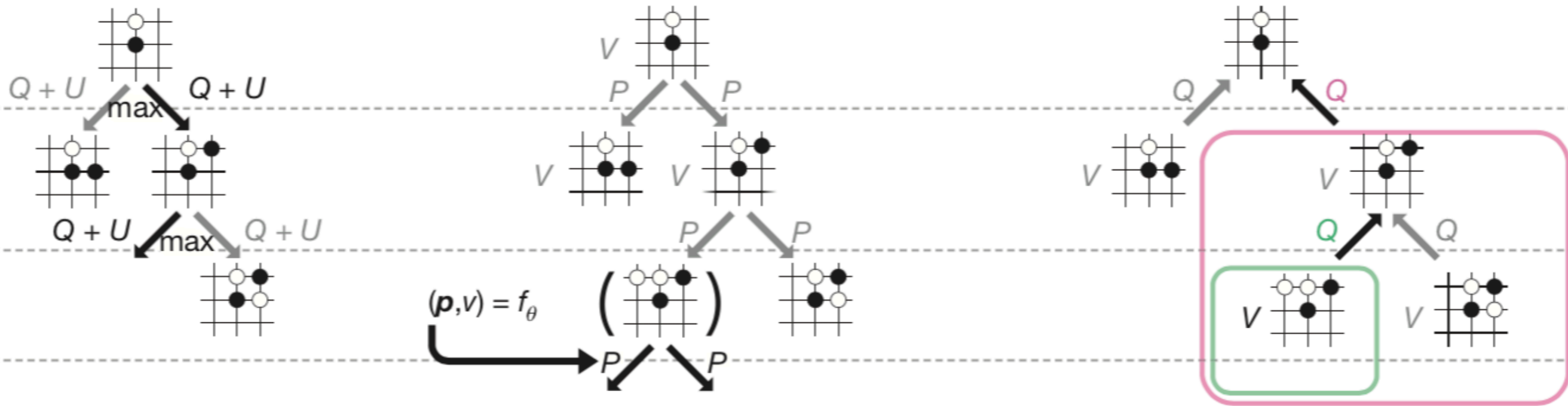
# AlphaGo

**a** Select

**b** Expand and evaluate

**c** Backup

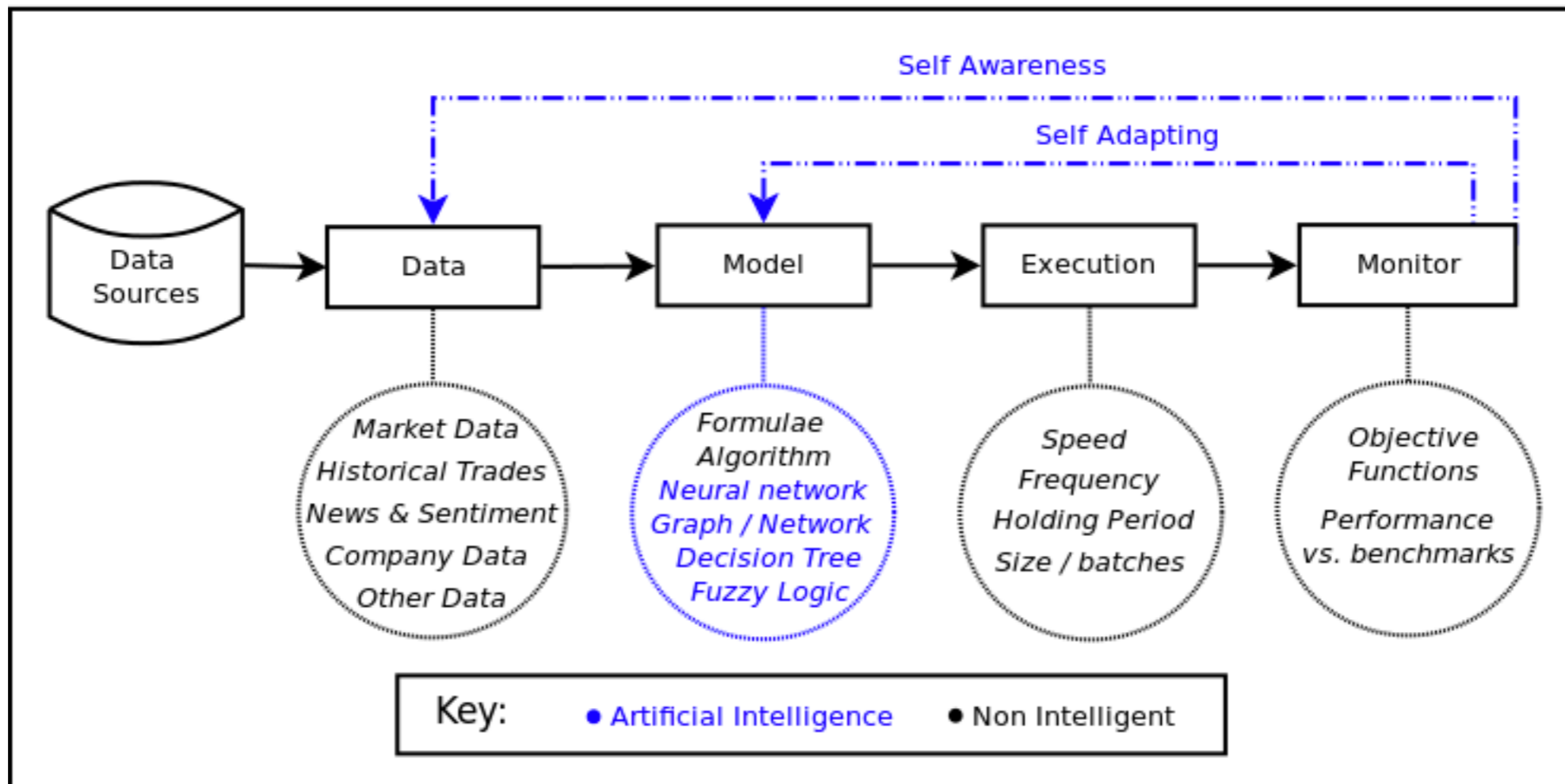
Repeat



<https://hackernoon.com/the-3-tricks-that-made-alphago-zero-work-f3d47b6686ef>

# Algorithmic trading

Conceptual Model of Algorithmic Trading



<http://www.turingfinance.com/wp-content/uploads/2013/11/Algorithmic-Trading-Systems-Conceptual.png>

# Conclusion

- Most statistical models are represented by probability models with certain unknown parameters
- Probability models are usually specified by human
- Unknown parameters are estimated (trained) from data



- Does the model good enough?
  - goodness of fit
- Which model is better for a specific data science problem?
  - no free lunch theorem
  - model selection (cross-validation, information criteria, LASSO, etc.)
  - model averaging (ensemble learning)