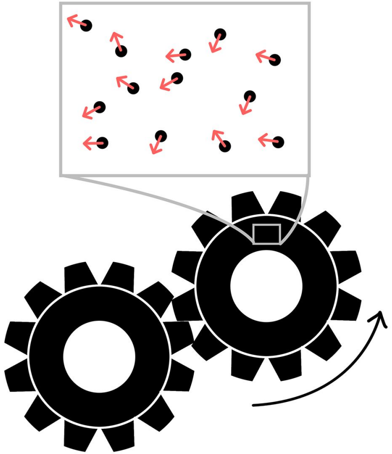
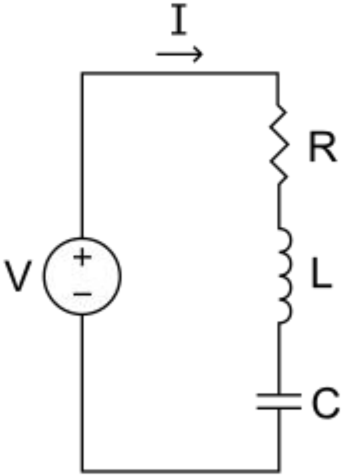
Abstraction = Information at a Distance

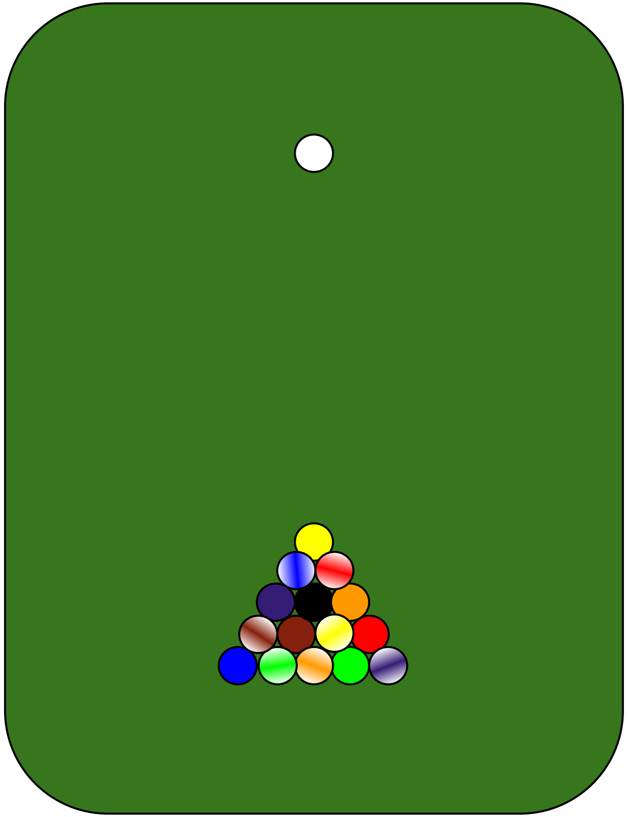
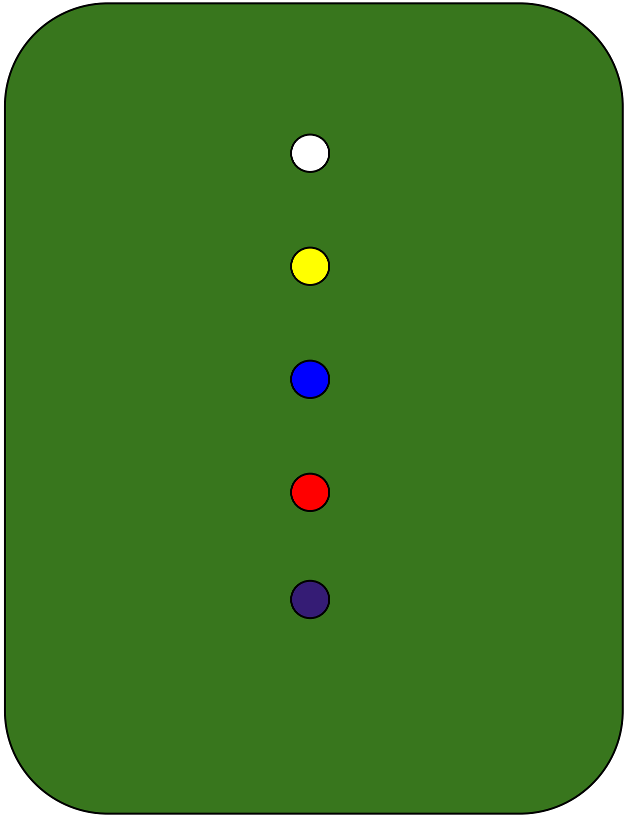
What Are We Talking About?

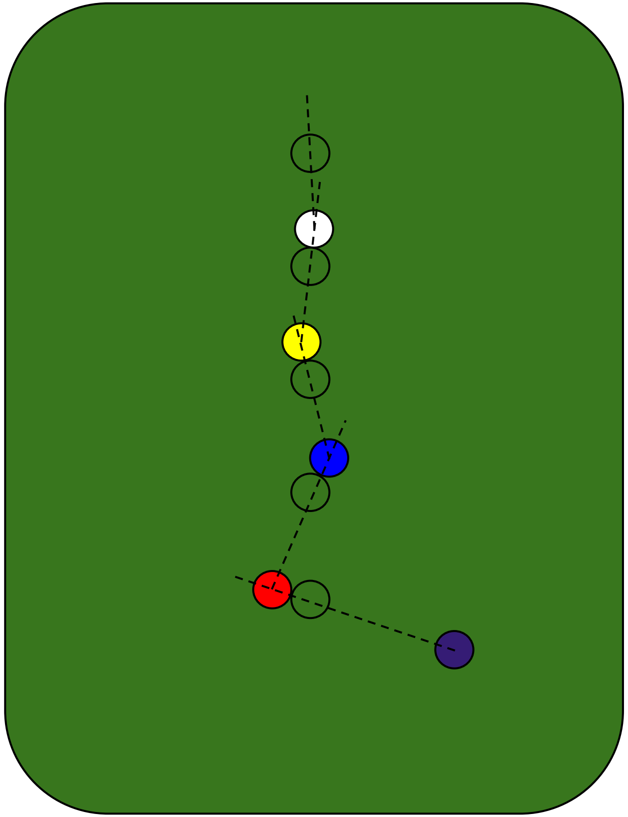




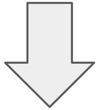
Outline

- Two stat-mech-flavored physical examples: billiards, electronics
- An everyday-flavored physical example: pencil
- General formulation (the math part)
- Science in a high-dimensional world: gears-level models
- Language in a high-dimensional world: clustering
- Natural abstraction hypothesis

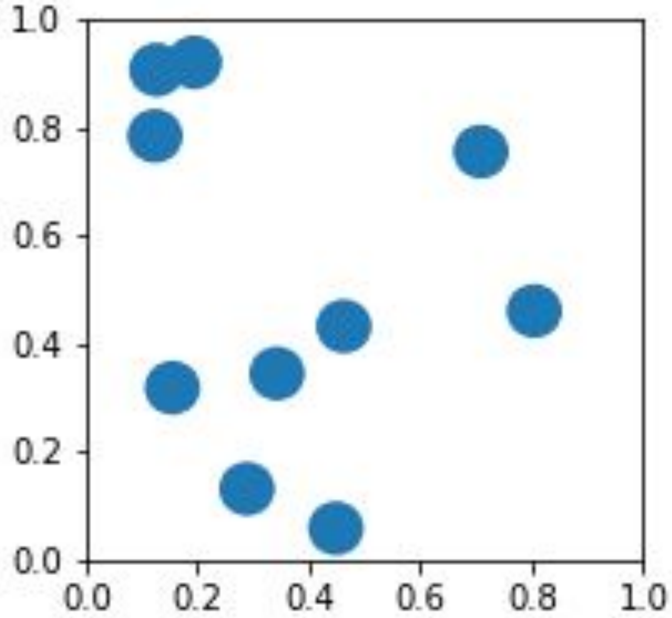





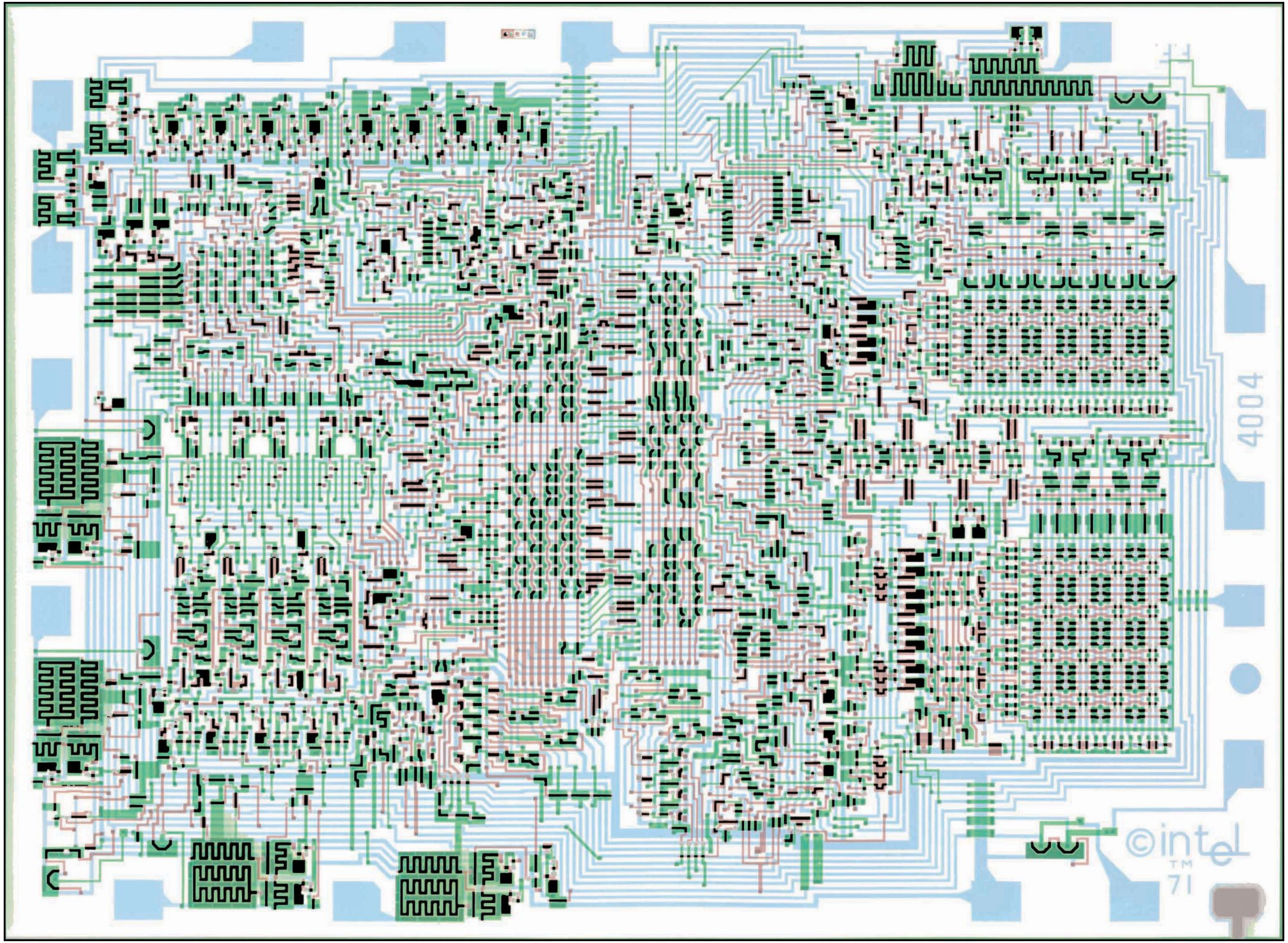
[0.50996900, 0.57680615, 0.97666898, …]

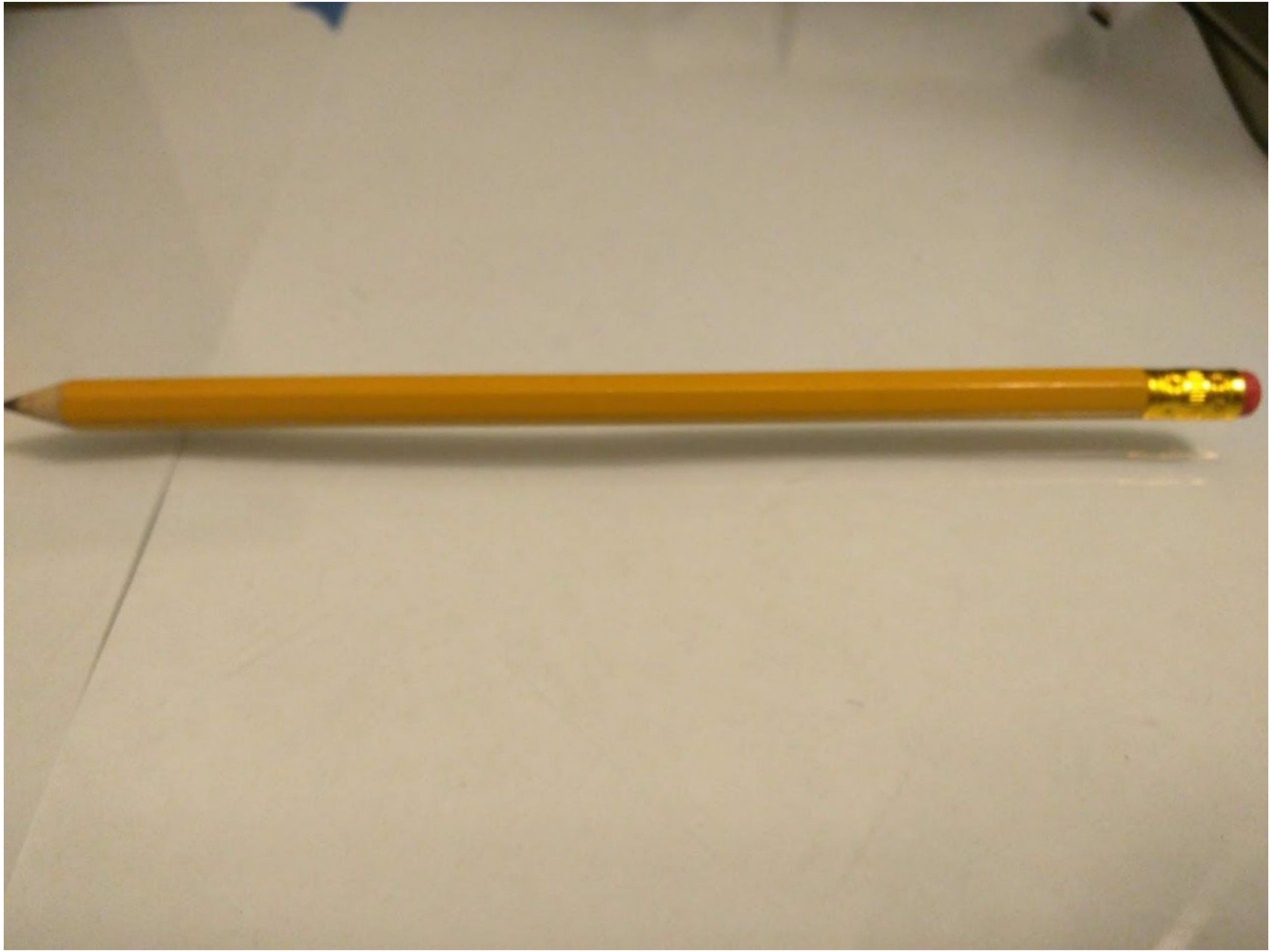


[0.22916602, 0.38694954, 0.98077806, …]

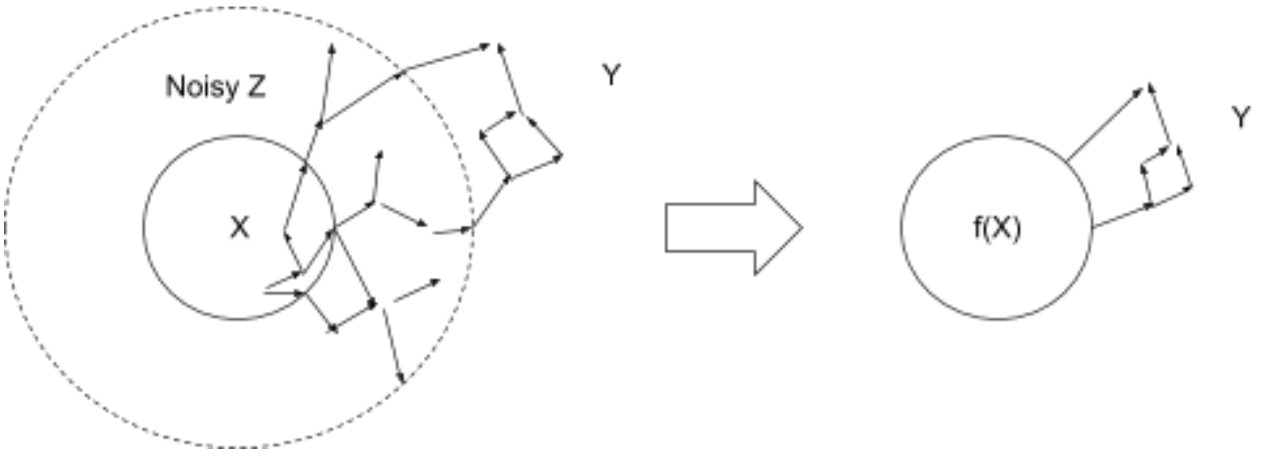


$$PV = nRT$$


4004
©intel
TM
71



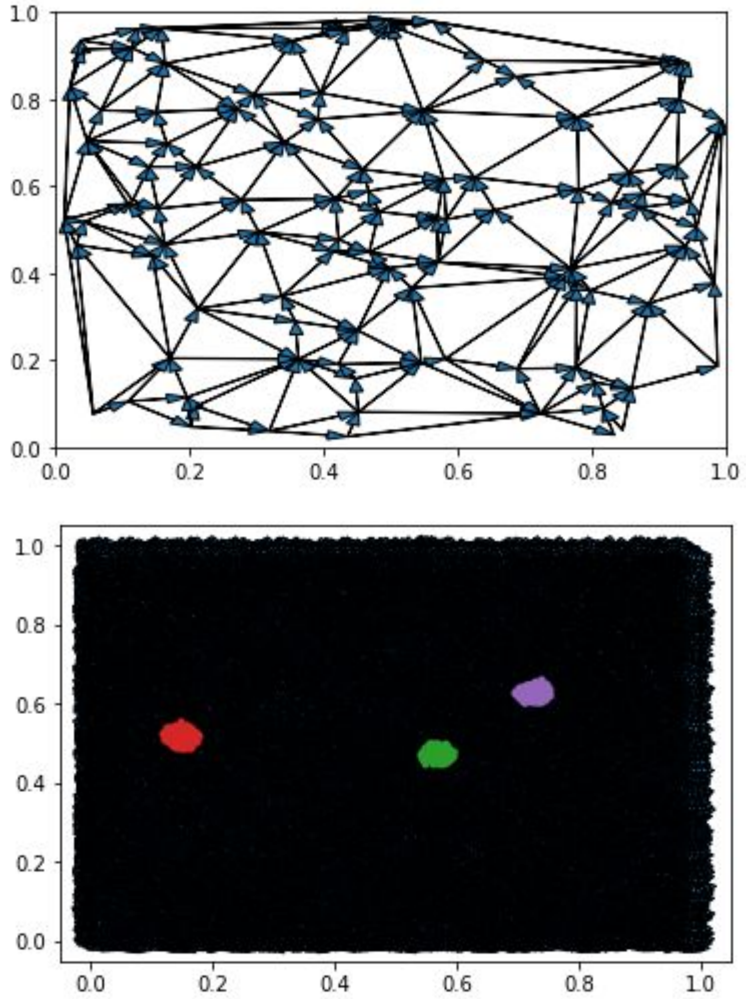
Formalization



Formula:

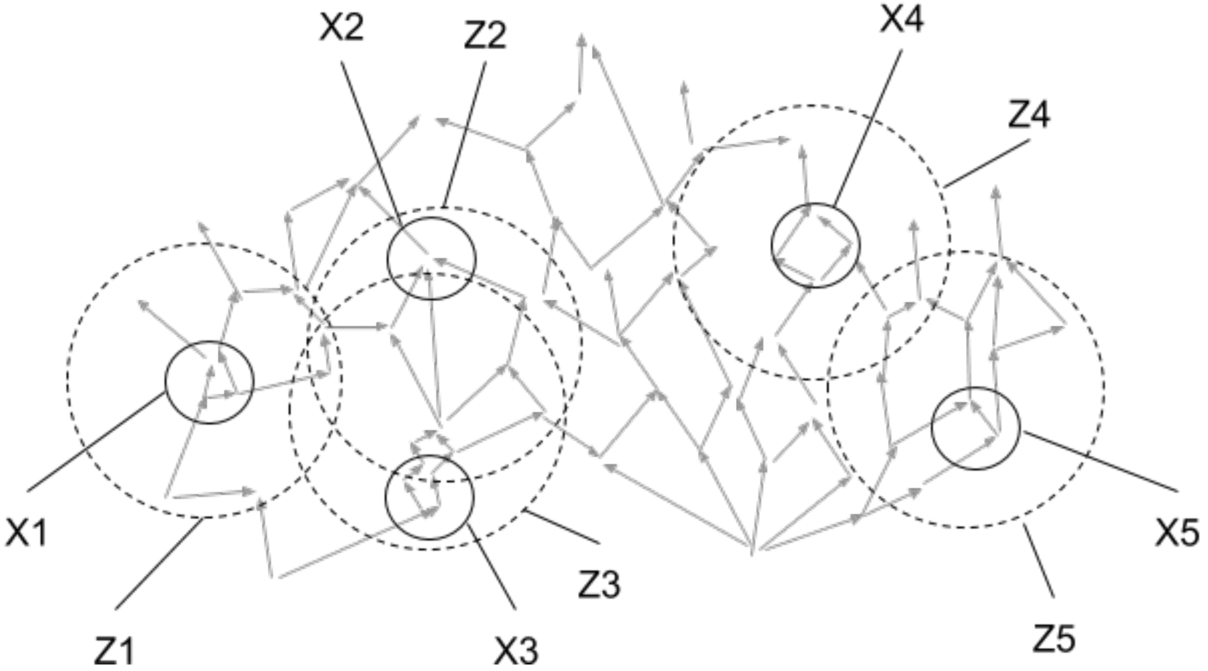
$$P[Y|X] = P[Y|f(X)]$$

… for any Y “far away from” X

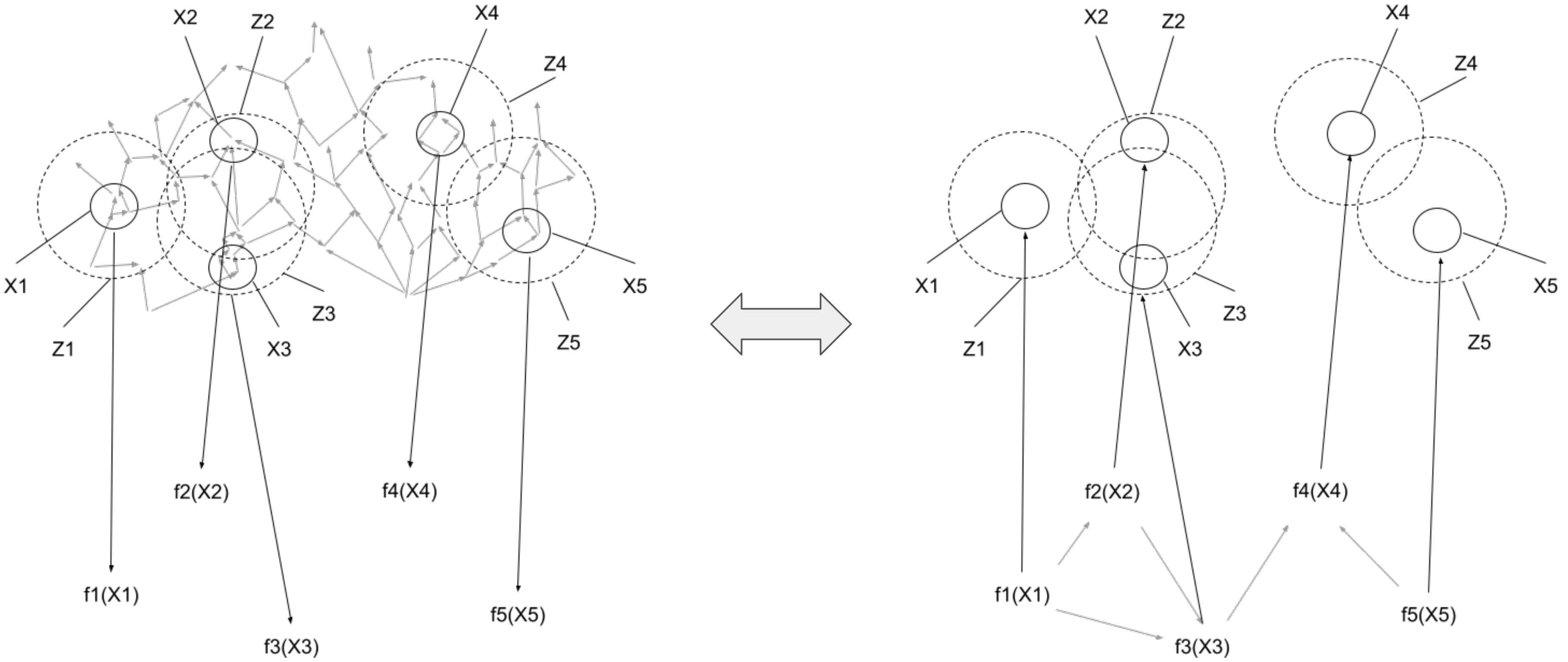


Covariance singular values:
[5.98e+05, 1.21e+02, 1.91e-01, 1.03e-03, 2.01e-04, ...]

System View



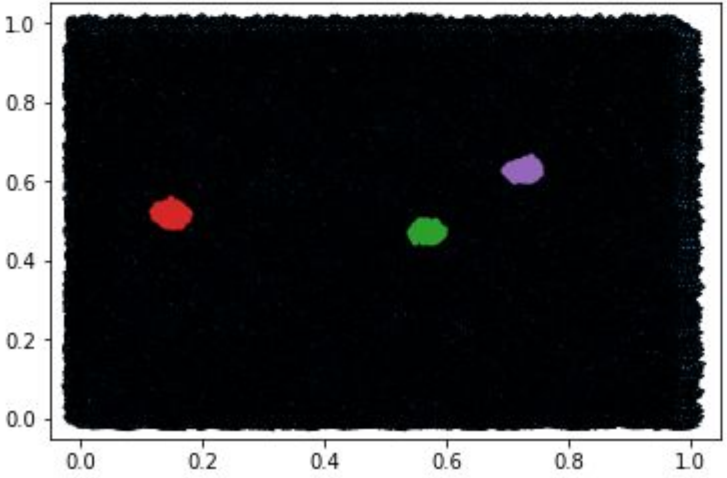
Each variable X_i has a set Z_i of variables which are “nearby”.



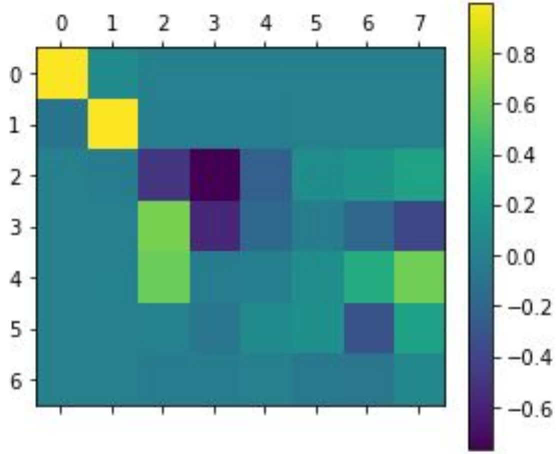
Formula:

$$P[X_S^L, X_S^H] = P[X_S^H] \prod_{i \in S} P[X_i^L | X_i^H] = P[X_S^L] \prod_{i \in S} P[X_i^H | X_i^L]$$

… so long as all variables in S are “far apart”.

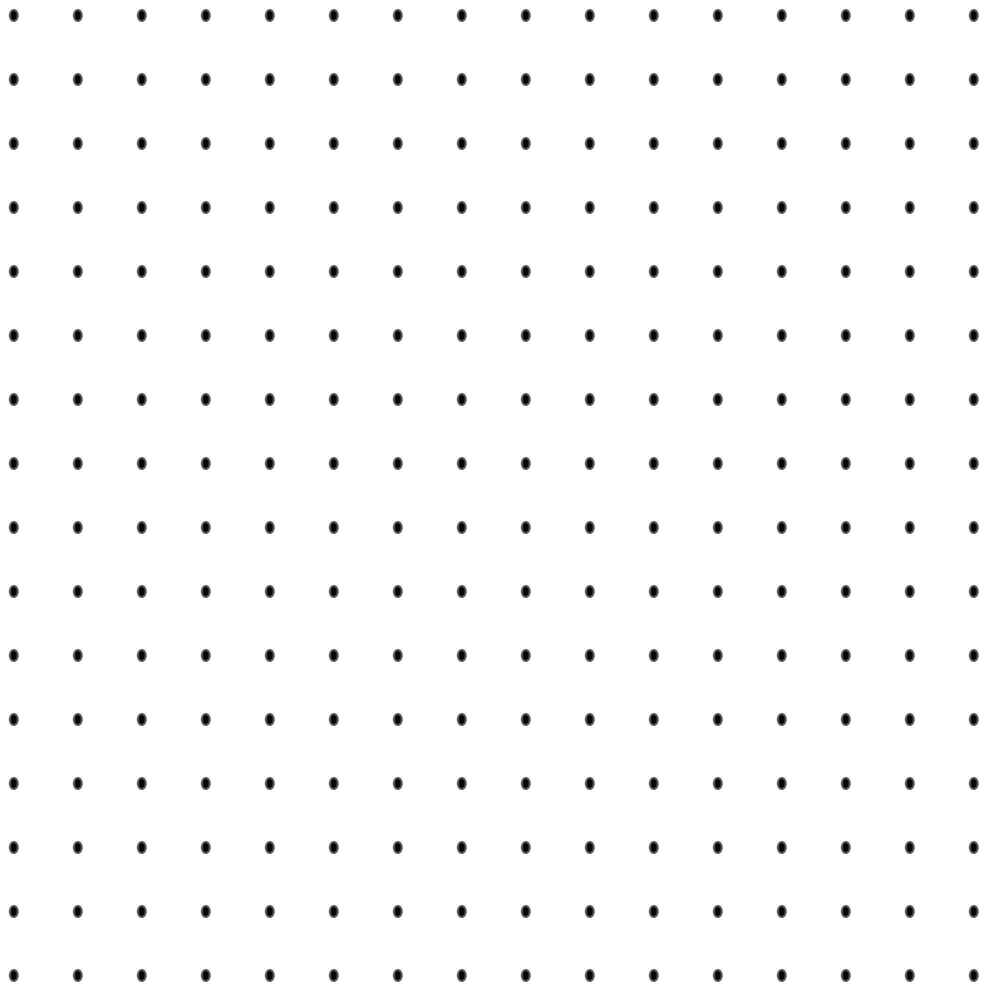


1-2 vs 2-3 covariance singular vectors



Application: Science in a High-Dimensional World


Calvin and Hobbes
Any of billions of variables could change this sled's trajectory.
A stray gust of wind or mound of snow, a fallen stick or accidental shift of weight, could change our course.
Sure, but billions?
Think of all the vibrations of all the atoms in the snow around us.
And what if modern science missed something?
The motions of stars, the flap of a far-away butterfly's wings... any of them could provide a little push which nobody noticed before, a signal in the "noise".
How could we ever hope to predict it?
Well, every time we do this, we seem to crash at exactly the same spot.
I guess those billions of variables didn't change our trajectory after all.
12/29



$$\frac{D\mathbf{u}}{Dt} = \frac{1}{\rho}\nabla \cdot \boldsymbol{\sigma} + \mathbf{g}$$

$X1 = a\, X2^{2}$

$\sin(X1+X2) = c$

$X1 \cdot X2^{2}/(X3-X4) = X5^{X6} + \tanh(X7/X8) - \dots$

$e^{X1} - X2^{3} = 0$

$X1 - 6\, X2 + \frac{1}{2}\, X3 + 12\, X4 + \dots = 0$

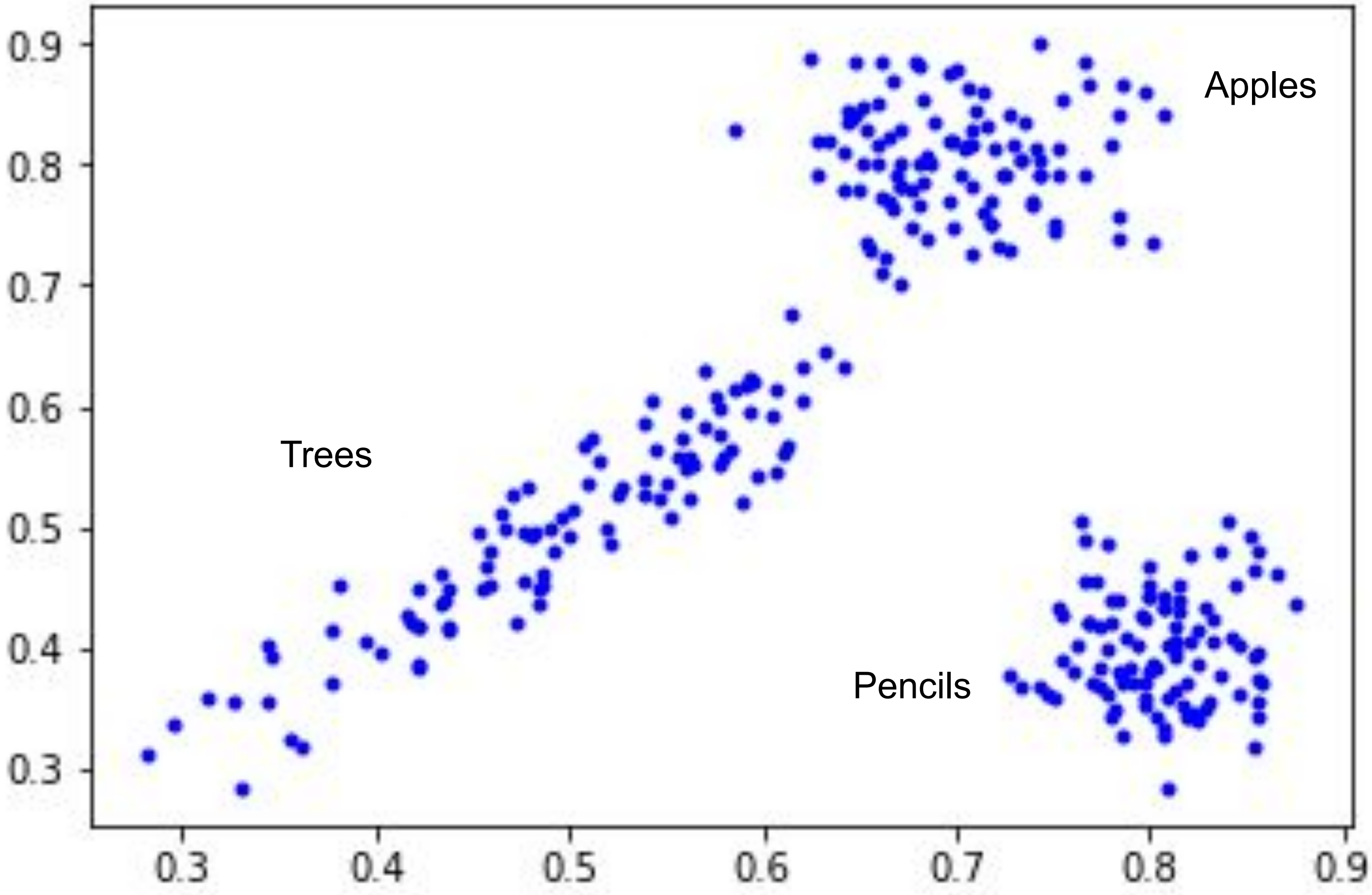
$X1 \ln(X1) + X2 \ln(X2) + e^{X3} \ln(X4) = X5 \cdot X6 + X7 + \dots$

$X1^{(X2 \cdot X3 - X4 \cdot X5)}/(3\, X6 - \dots) = 0$

$X1 = X2^{2} \cdot J2(X3 - X4/X5) + \dots$

$\mathrm{sqrt}(X1 + X2/X3) + 4\, X4 = \ln(e^{X5} + e^{X6} + 1) \cdot (X7 + \dots)$

Application: Language in a High-Dimensional World


Apples
Trees
Pencils
0.3
0.4
0.5
0.6
0.7
0.8
0.9

Natural Abstraction Hypothesis

Summary

- Abstraction, in day-to-day practice, usually involves summarizing the information from some chunk of the world which is relevant “far away”.
- Empirically, it turns out that we can pick the chunks so that the summaries are low-dimensional (compared to our high-dimensional world).
- Something like this has to be true pretty often in order for science to work the way it does.
- This also provides a conceptually-nice model for language foundations.
- The Natural Abstraction Hypothesis says that most human concepts work this way, and that a wide range of cognitive architectures converge to approximately the same abstract concepts.

To read more, look for Testing the Natural Abstraction Hypothesis on lesswrong.com; it contains many links to related posts.