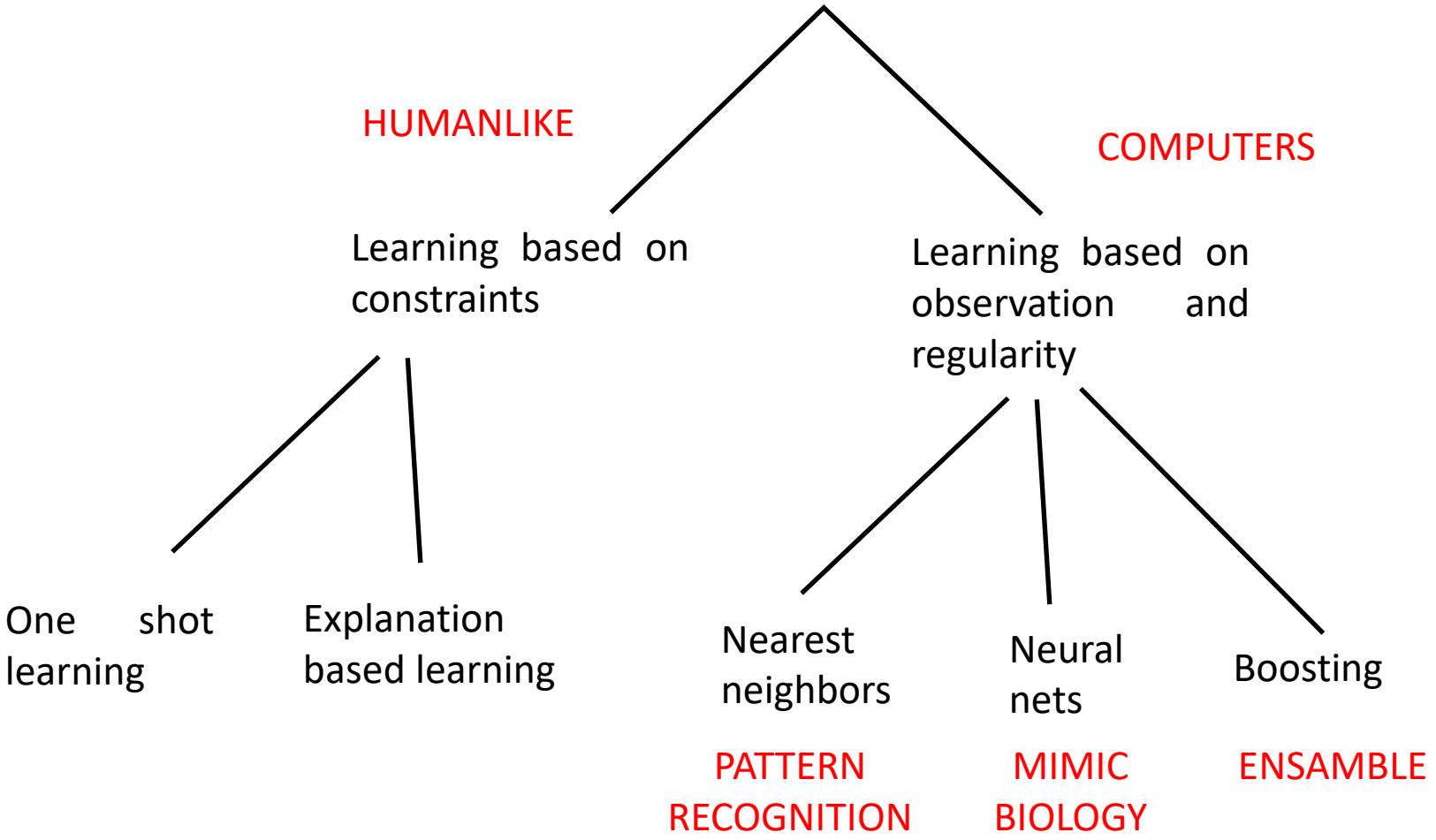
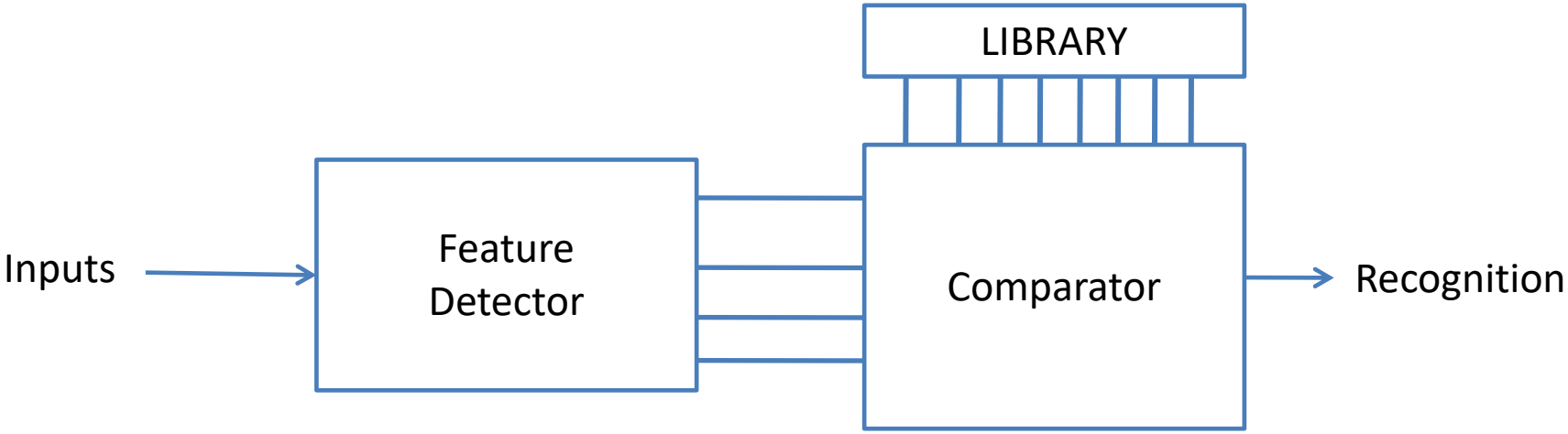
Artificial Intelligence Fundamentals

Learning: Nearest Neighbors

Learning

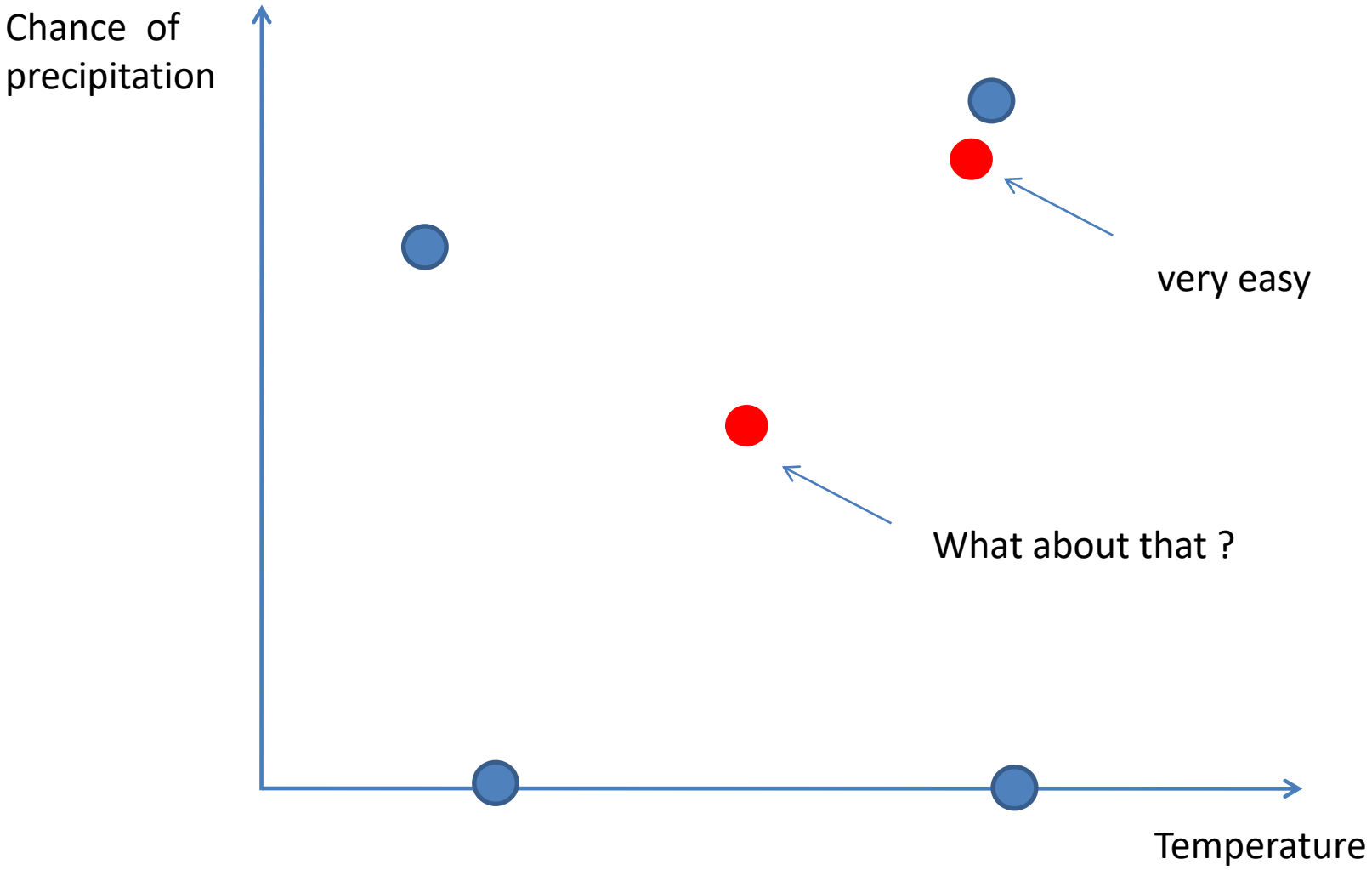

HUMANLIKE
COMPUTERS
Learning based on constraints
Learning based on observation and regularity
One shot learning
Explanation based learning
Nearest neighbors
Neural nets
Boosting
PATTERN RECOGNITION
MIMIC BIOLOGY
ENSAMBLE

Nearest neighbors

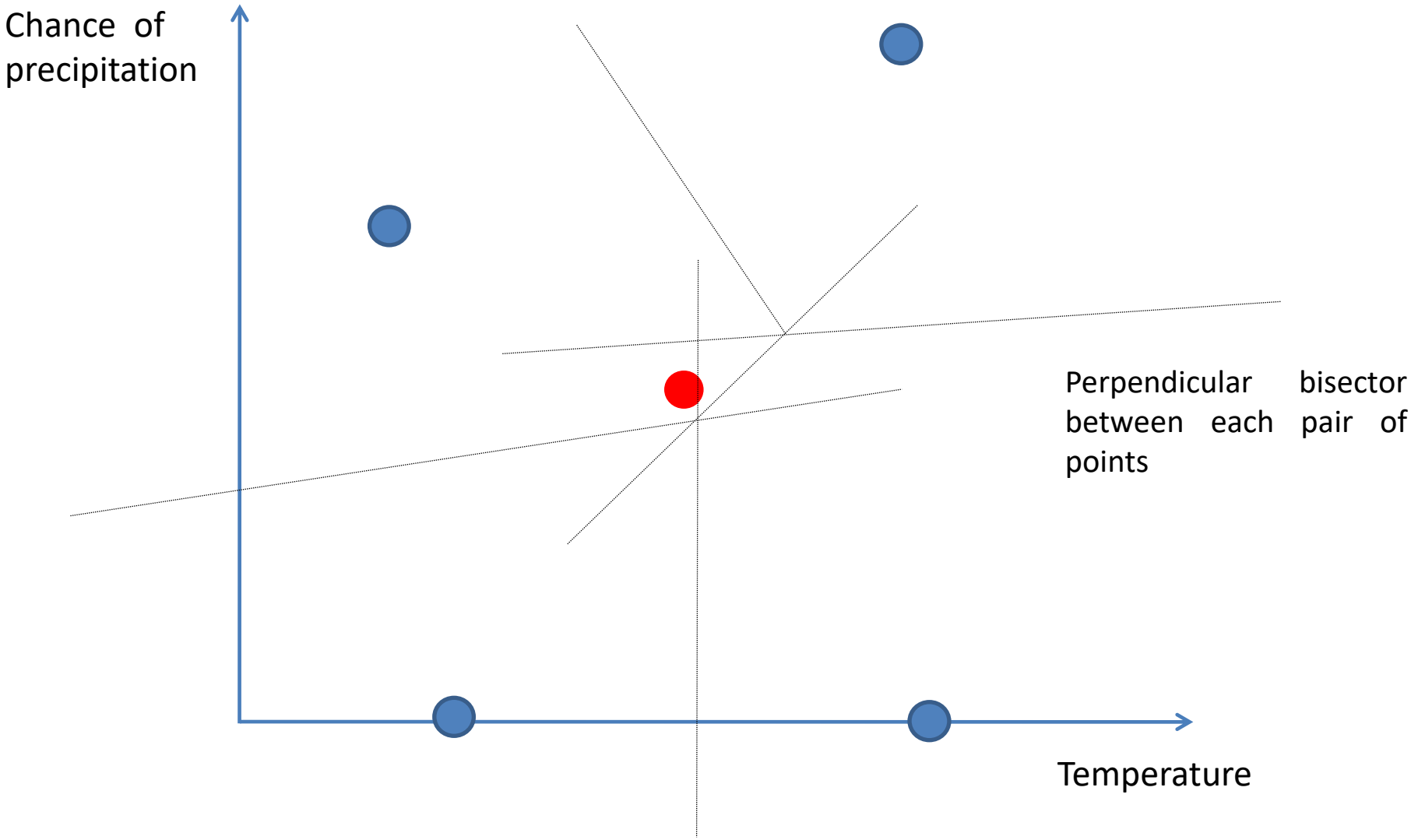


- Feature detector – extracts a vector of features from the inputs and for each feature compute a value
- LIBRARY – stores multiple feature vectors, one feature vector for one possibility
- Comparator – compare the feature vector with all the possibilities and return the nearest possibility -> recognition

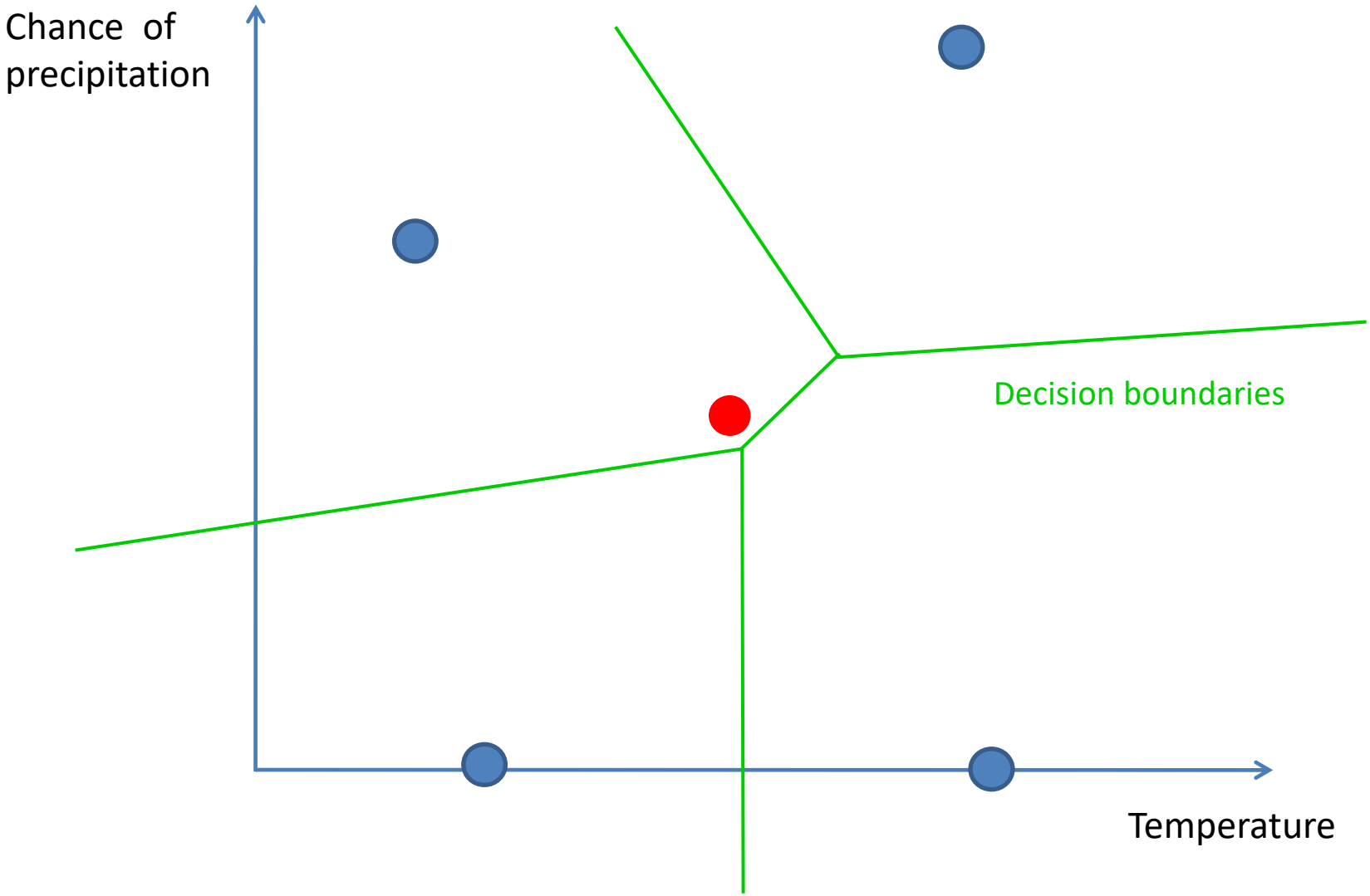
How we dress today?



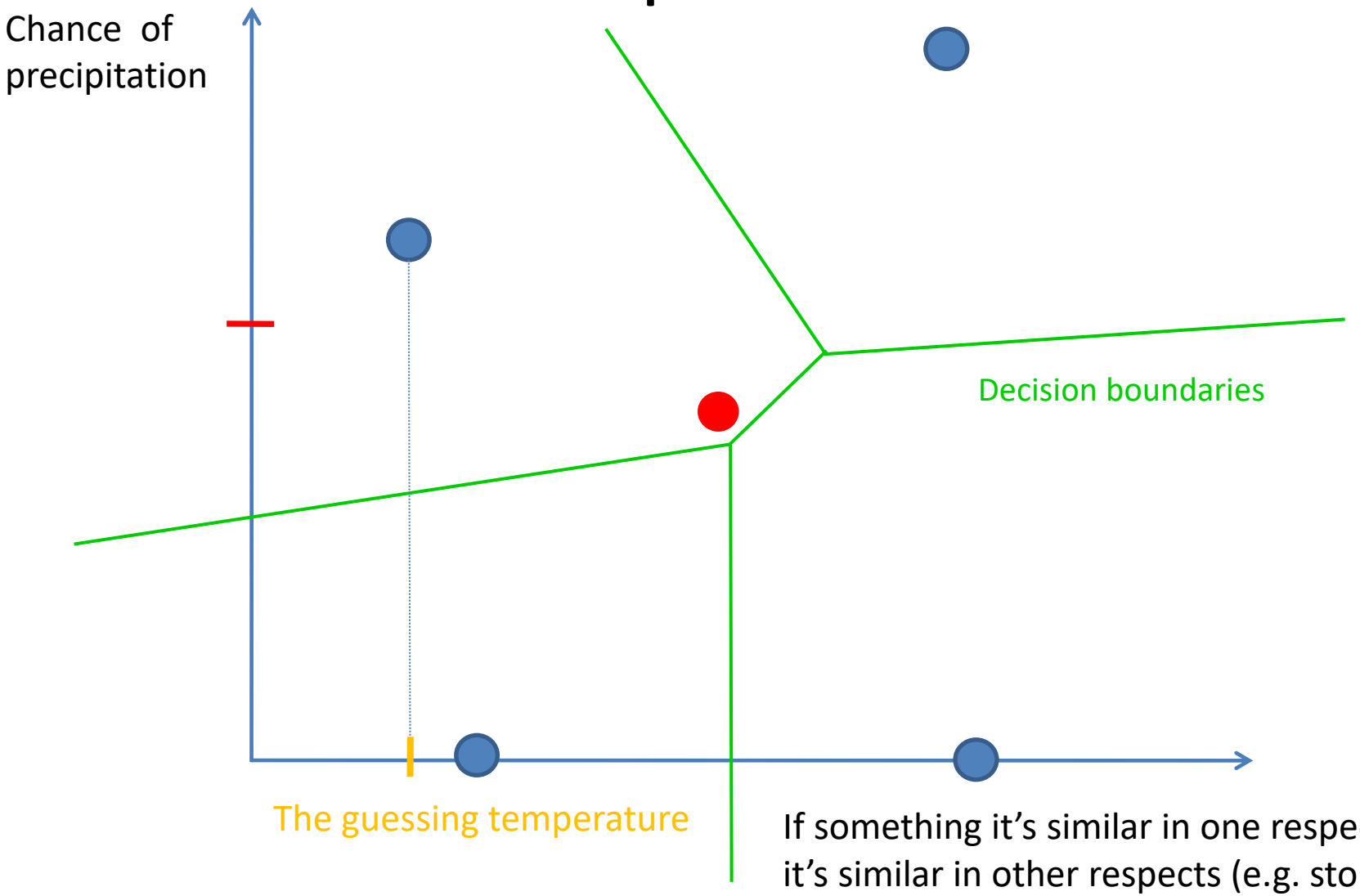
How we dress today?



How we dress today?

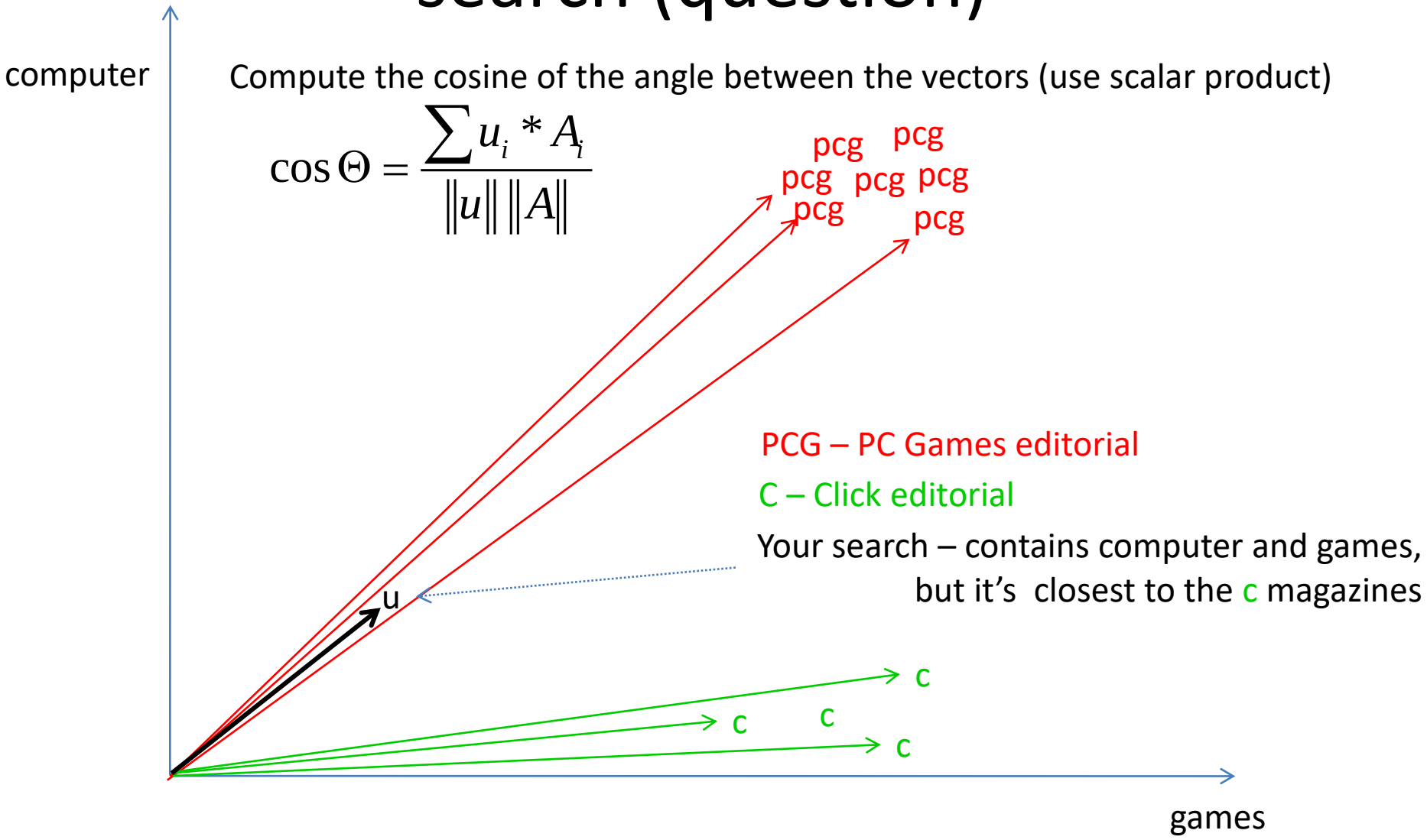


What if I measure only the temperature?

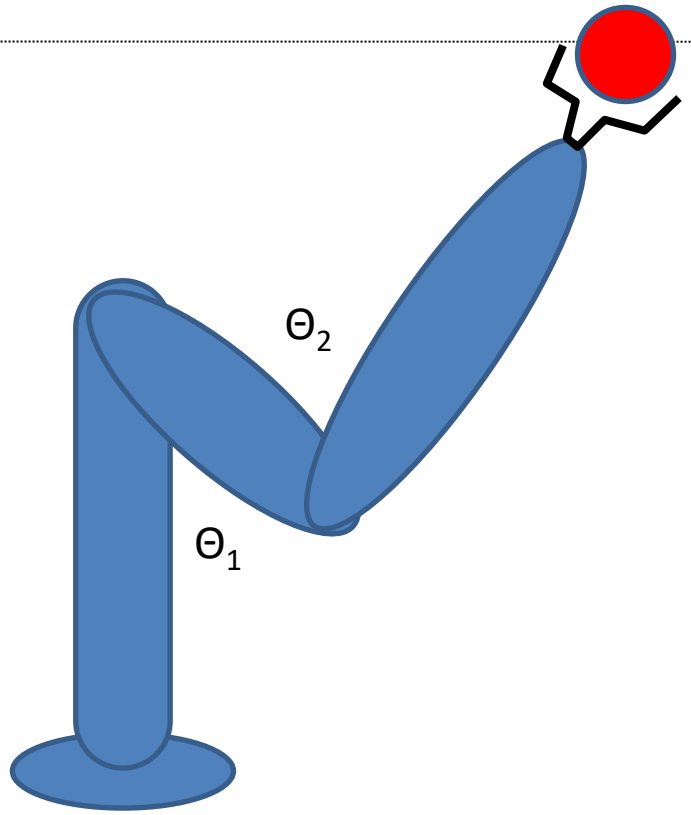


If something it's similar in one respect, it's similar in other respects (e.g. stories, legal cases, medical cases, etc.)

Finding magazines relevant to your search (question)



Robotic arm control-move the ball

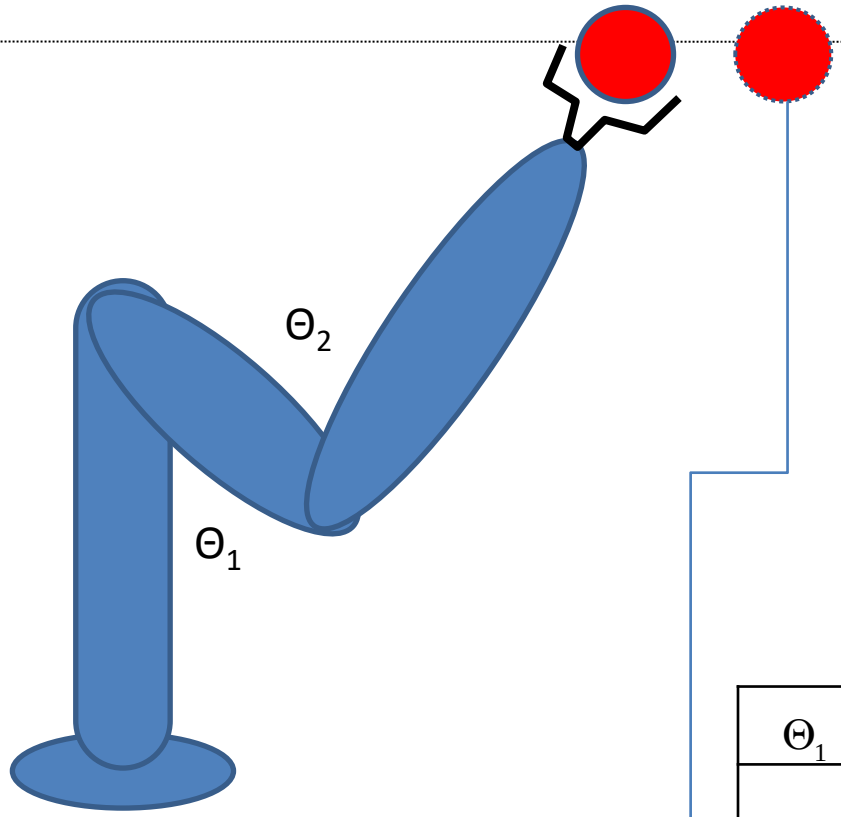


- Cinematic problem – translating the x coordinate of the ball into the Θ_1 and Θ_2 space
- Mechanical problem - Solve the equations – but there are very complicated

---

- Even so it doesn't work – there are frictions, precise of the measurements, etc.

We must fill the table



- For each position we have a corresponding row (the most similar one) in the table

torque

Θ_1	Θ_2	$\dot{\Theta}_1$	$\dot{\Theta}_2$	$\ddot{\Theta}_1$	$\ddot{\Theta}_2$	T_1	T_2

K Nearest Neighbors

- 1-NN
  - Given an unknown point, pick the closest 1 neighbor by some distance measure.
  - Class of unknown is the 1-nearest neighbor's label.
- k-NN
  - Given an unknown, pick the k closest neighbors by some distance function.
  - Class of unknown is the **mode** of the k-nearest neighbor's labels.
  - k is usually an odd number to facilitate tie breaking.

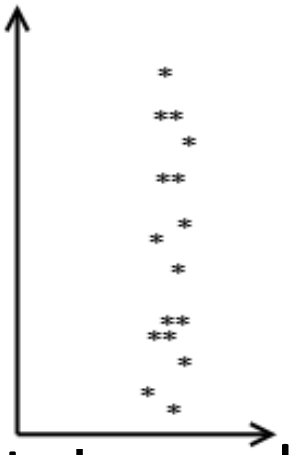
How to draw 1-NN decision boundaries

- Decision boundaries, lines on which it is equally likely to be in any of the classes.
  1. Examine the region where you think decision boundaries should occur.
  2. Find oppositely labeled points (+/-)
  3. Draw bisectors. (use pencil)
  4. Extend and join all bisectors. Erase extraneously extended lines.
  5. Remember to draw boundaries to the edge of the graph and indicate it with arrows! (a very common mistake)
  6. Your 1-NN boundaries generally should have sharp edges and corners (otherwise, you are doing something wrong or drawing boundaries for a higher k-nn.)

Problems

- SPREAD problem
  - Normalize the data

$$\sigma_x^2 = \frac{1}{N}\sum(x - \bar{x})^2 \quad , x' = \frac{x}{\sigma_x}$$



- WHAT MATTERS problem, the result depends only on *x* (not *y*)
  - The answer will be wrong

- NO PROBLEM problem, the answer doesn't depend the data at all

Related resources

- http://ocw.mit.edu/courses/electrical-engineering-and-computer-science/6-034-artificial-intelligence-fall-2010/exams/MIT6_034F10_quiz2_2007.pdf

Readings

- Artificial Intelligence (3rd Edition), Patrick Winston, Chapter 12