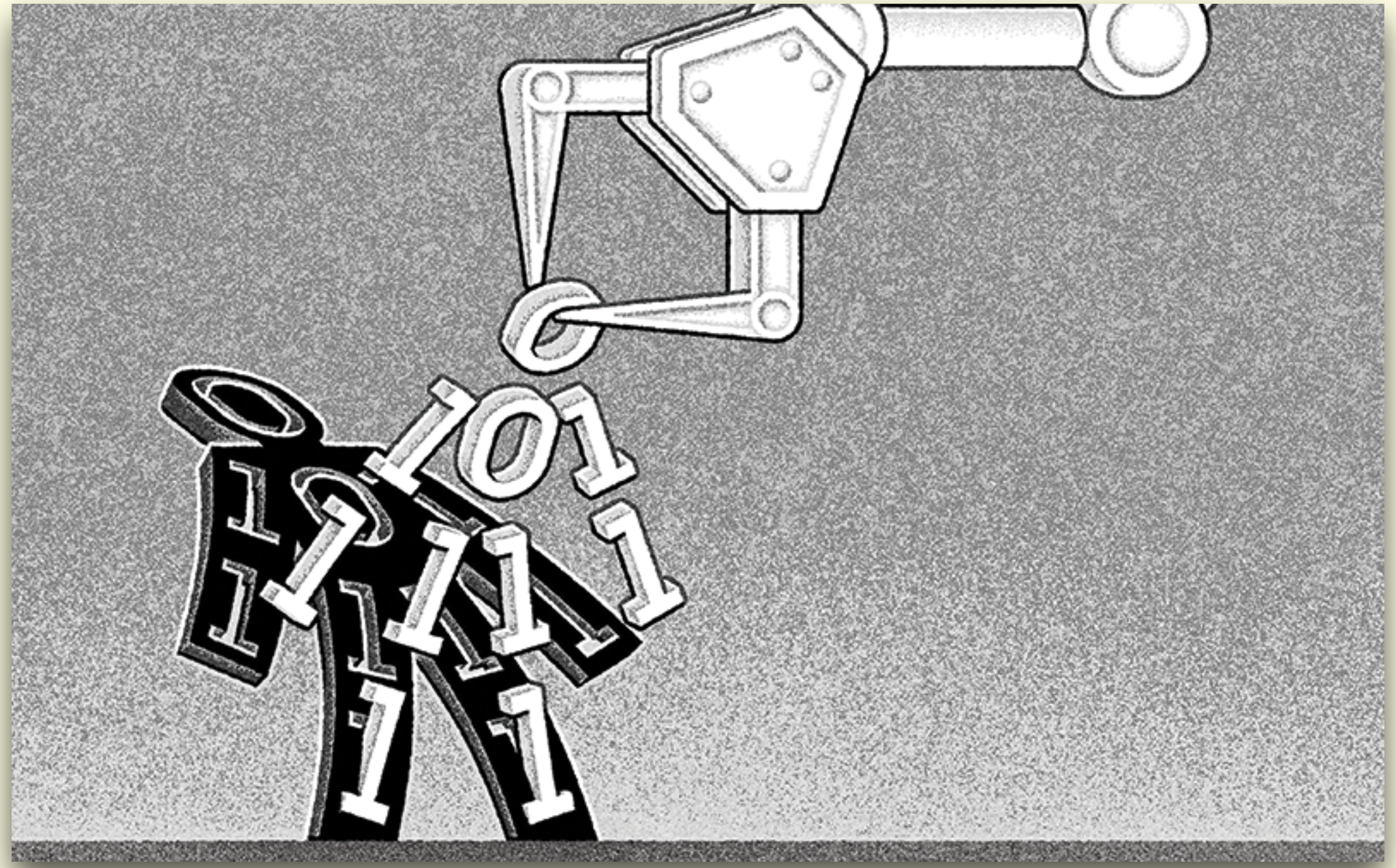
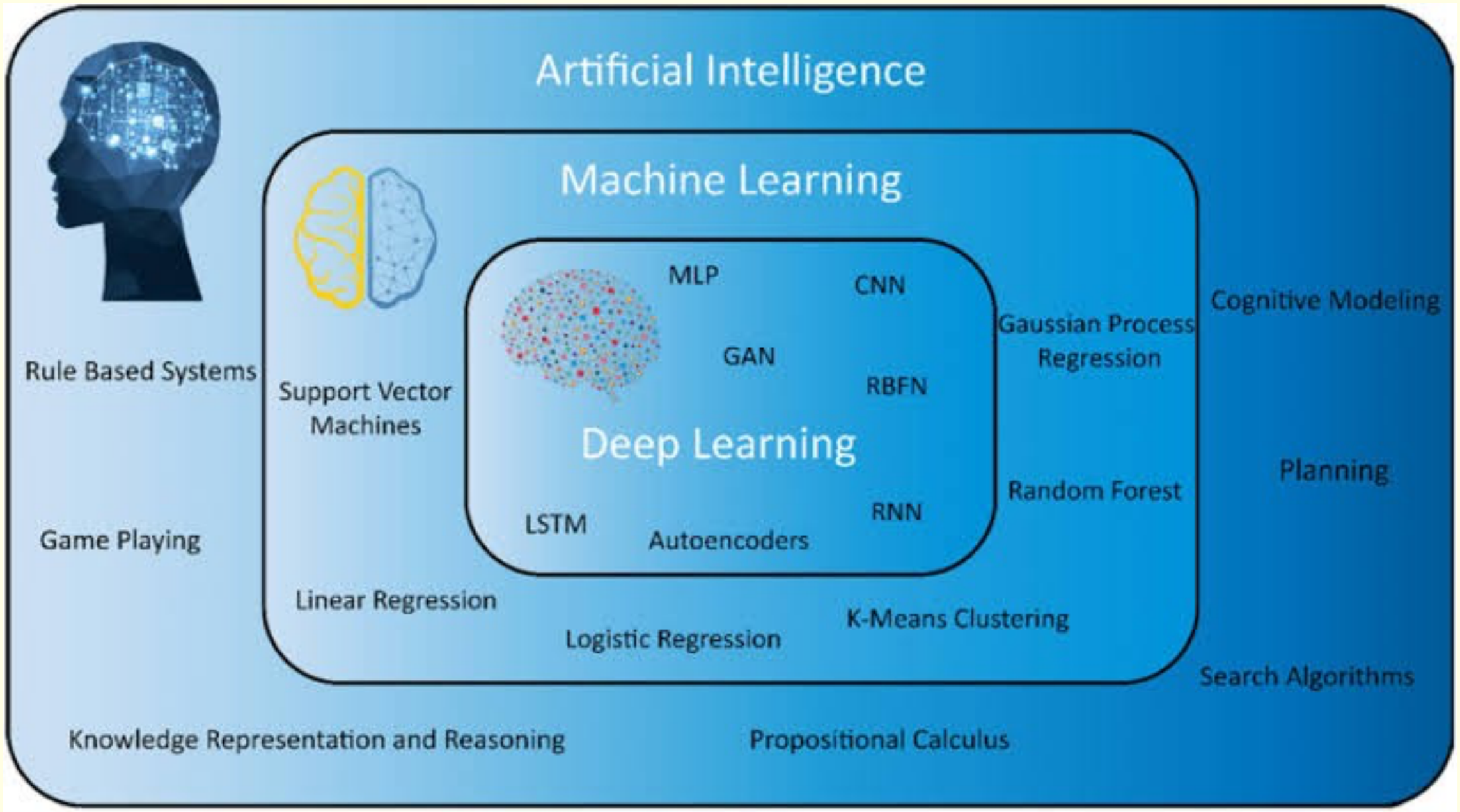


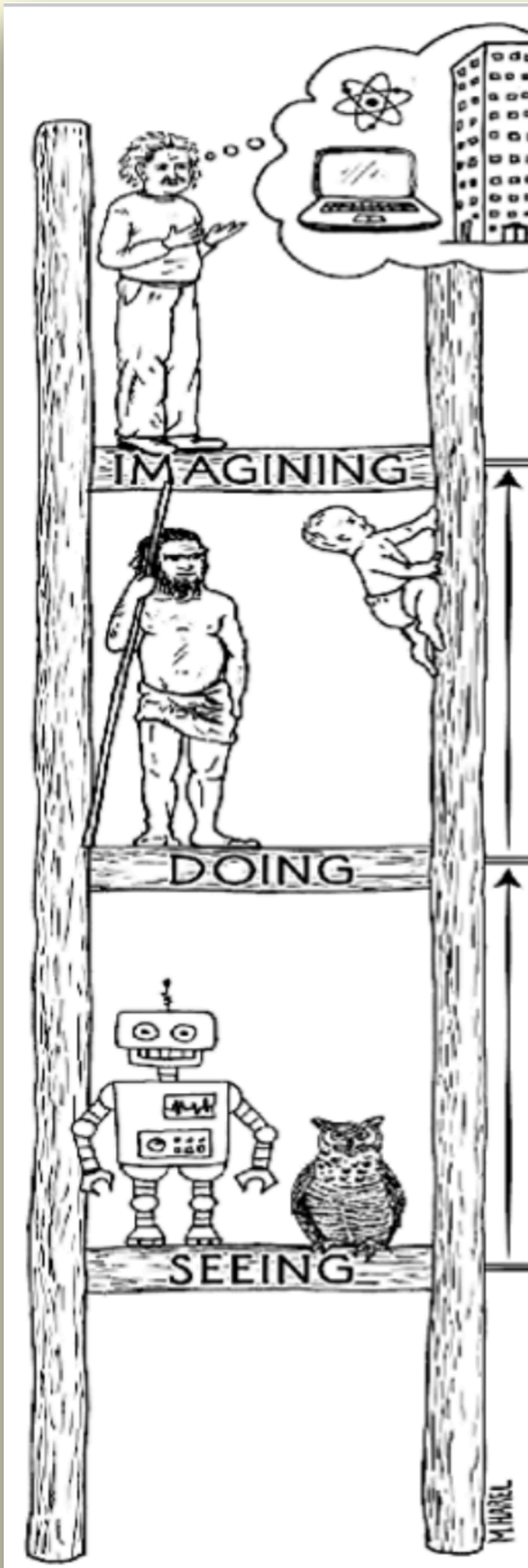
https://www.theactuary.com/features/2020/08/05/machine-learning-deep-end

AI and ML - Well, it is all about statistical and probabilistic ideas...

An introduction

Dr. Pat Ryser-Welch / Oct 2023


Artificial Intelligence
Machine Learning
Deep Learning
MLP
CNN
GAN
RBFN
LSTM
Autoencoders
RNN
Rule Based Systems
Support Vector Machines
Gaussian Process Regression
Random Forest
Cognitive Modeling
Planning
Game Playing
Linear Regression
Logistic Regression
K-Means Clustering
Search Algorithms
Knowledge Representation and Reasoning
Propositional Calculus



3. COUNTERFACTUALS

ACTIVITY: Imagining, Retrospection, Understanding

QUESTIONS: *What if I had done ...? Why?*
(Was it X that caused Y? What if X had not occurred? What if I had acted differently?)

EXAMPLES: Was it the aspirin that stopped my headache? Would Kennedy be alive if Oswald had not killed him? What if I had not smoked for the last 2 years?

2. INTERVENTION

ACTIVITY: Doing, Intervening

QUESTIONS: *What if I do ...? How?*
(What would Y be if I do X?
How can I make Y happen?)

EXAMPLES: If I take aspirin, will my headache be cured?
What if we ban cigarettes?

1. ASSOCIATION

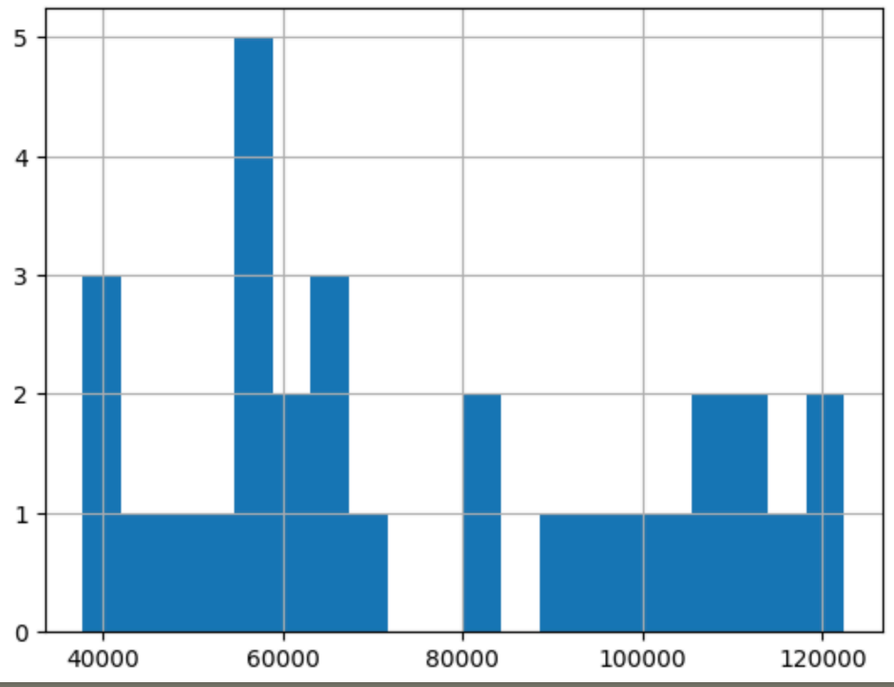
ACTIVITY: Seeing, Observing

QUESTIONS: *What if I see ...?*
(How are the variables related?
How would seeing X change my belief in Y?)

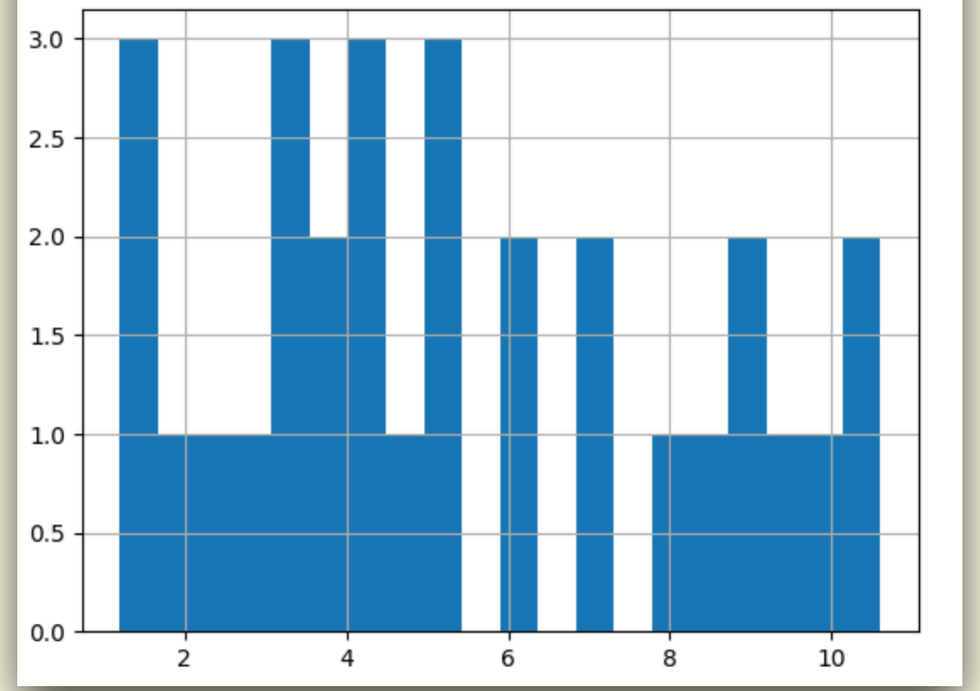
EXAMPLES: What does a symptom tell me about a disease?
What does a survey tell us about the election results?

A simple association: Years of experience and Salary

```
Out[141…  count        30.000000
          mean      76004.000000
          std       27414.429785
          min       37732.000000
          25%       56721.750000
          50%       65238.000000
          75%      100545.750000
          max      122392.000000
          Name: Salary, dtype: float64
```

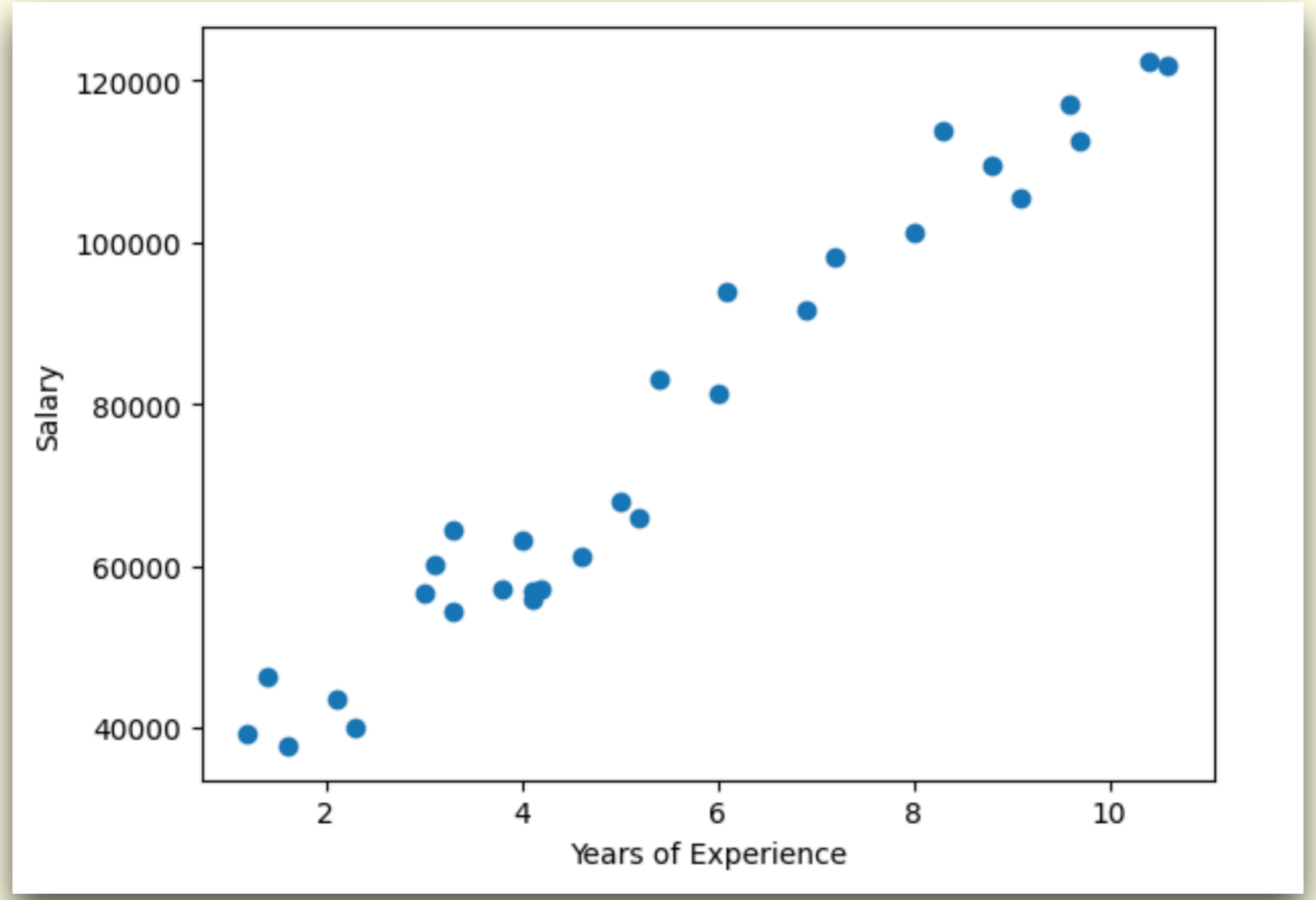



```
count    30.000000
mean      5.413333
std       2.837888
min       1.200000
25%       3.300000
50%       4.800000
75%       7.800000
max      10.600000
Name: YearsExperience, dtype: float64
```



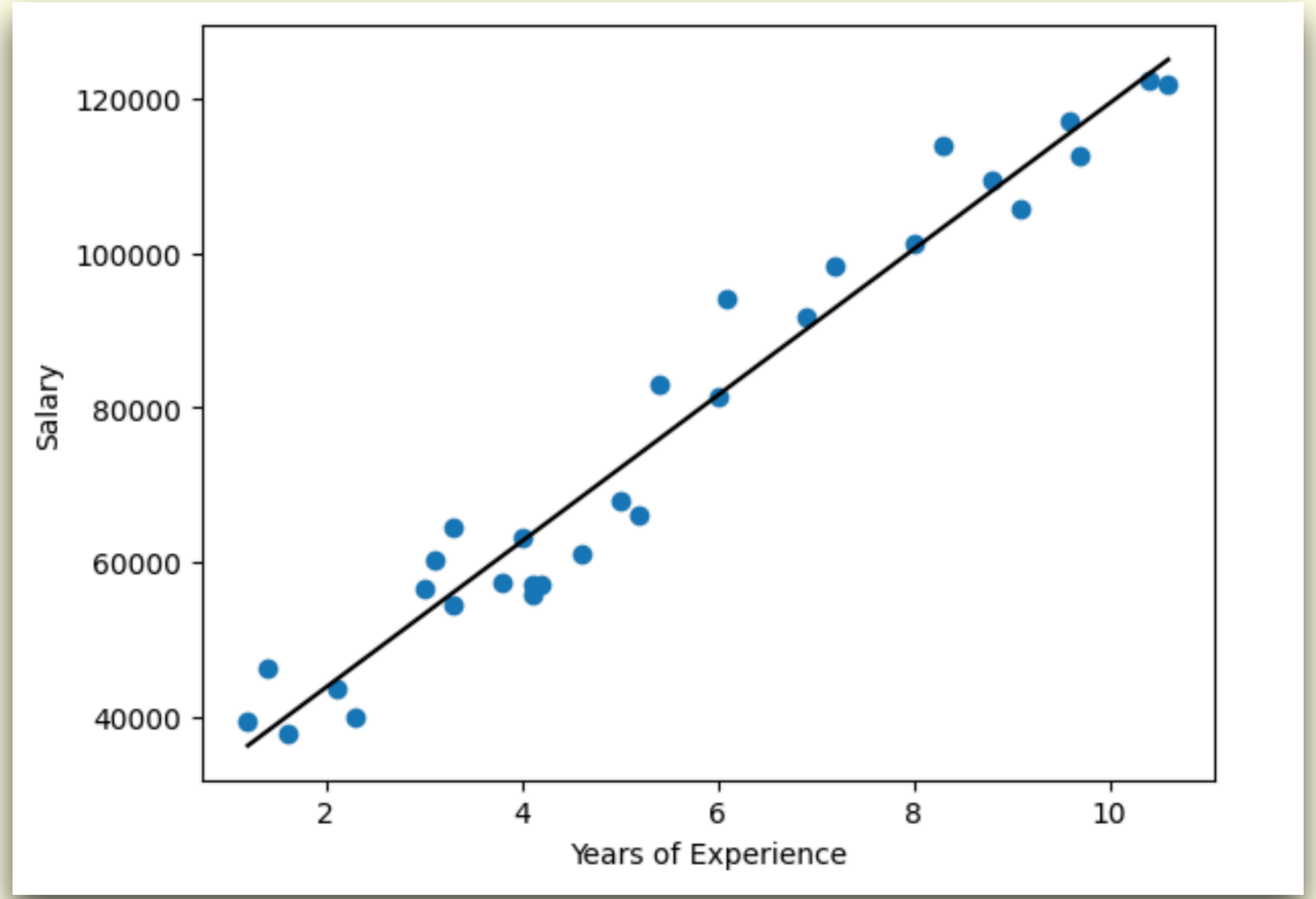
Years of Experience

In your opinion, what type of pattern may associate the year of experience to the salary?



Years of Experience

In your opinion, what type of pattern may associate the year of experience to the salary?



What did we do?

- We have a set of labelled examples - The observed salary.
- We have a set of other values that may be matching the labelled example.
- We constructed a model, based on a hypothesis.
- We use a sample that we surmised represented some properties about years of experience and salary.

$$\text{Model} = \text{data} + \text{error}$$

Metrics - error and accuracy

In your opinion, would you interpret this percentage error of predictions?

```
array([0.07047916])
```

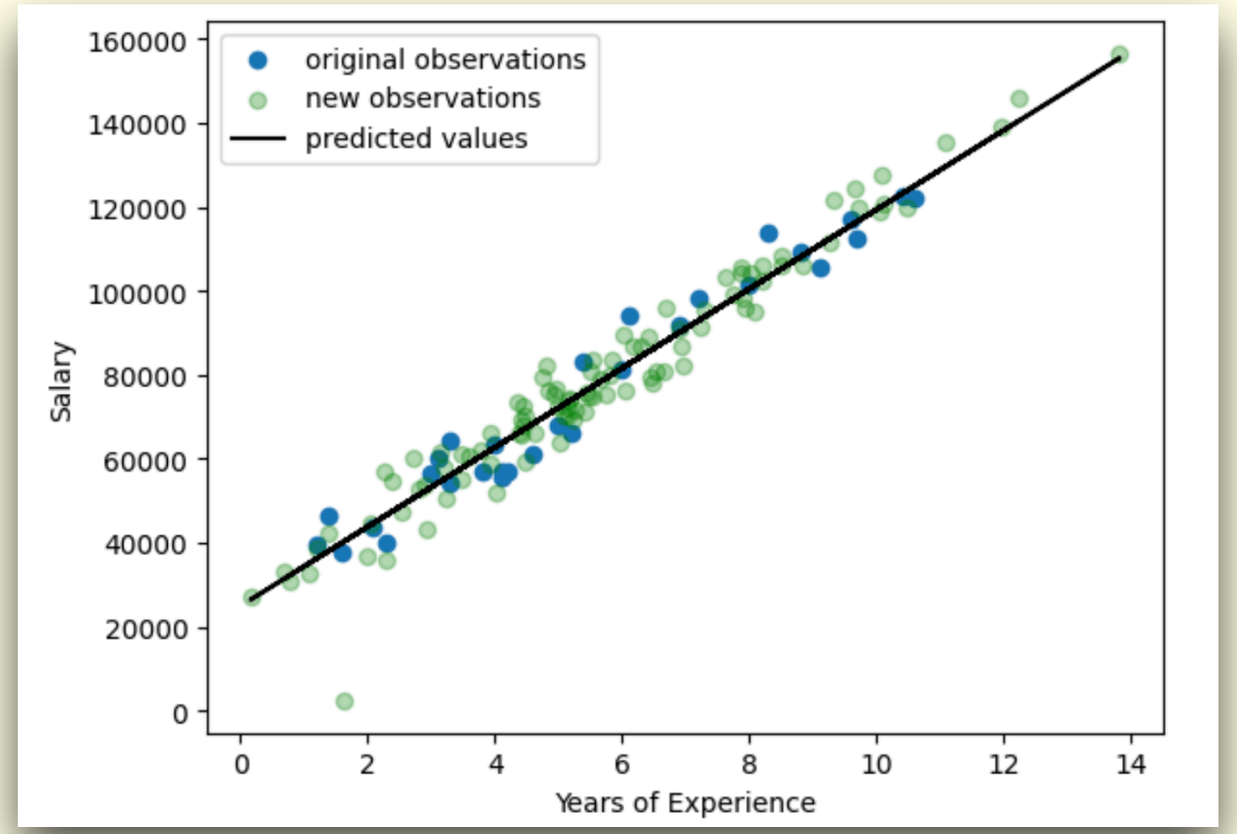
Metrics - error and accuracy

We use the mean absolute average error to evaluate the model. The model appears to be quite accurate. The error is less than 10% and, consequently, the accuracy is 90% or greater.

```
mape_sum = sum(abs(ys - Y_pred)/ys)
mape = (1/len(Xs)) * mape_sum
mape
```

```
array([0.07047916])
```

Is this model suitable for other observations?



Notebook

What is PAC learning?

Definition 1 *(PAC learnability) Let $\mathcal{C}$ be a class of boolean functions $f : \{0,1\}^n \rightarrow 0,1$. We say that $\mathcal{C}$ is PAC-learnable if there exists an algorithm $\mathcal{L}$ such that*
for every $f \in \mathcal{C}$
for any probability distribution $\mathcal{D}$
for any ϵ $(where 0 \leq \epsilon < \frac{1}{2})$
for any δ $(where 0 \leq \delta < 1$
algorithm $\mathcal{L}$ on input ϵ and δ and a set of random examples picked from any probability distribution $\mathcal{D}$ outputs at least with a probability $1 - \delta$, concept h such that $error(h, f) \leq \epsilon$.

A PAC learning algorithm refers to an algorithm that returns a hypothesis that is PAC.

Using formal methods, a minimum generalization error can be specified for a supervised learning task. The theorem can then be used to estimate the expected number of samples from the problem domain that would be required to determine whether a hypothesis was PAC or not. That is, it provides a way to estimate the number of samples required to find a PAC hypothesis.

> *The goal of the PAC framework is to understand how large a data set needs to be in order to give good generalization. It also gives bounds for the computational cost of learning …*

— Page 344, Pattern Recognition and Machine Learning, 2006.

Questions explored in computational learning theory might include:

- How do we know a model has a good approximation for the target function?
- What hypothesis space should be used?
- How do we know if we have a local or globally good solution?
- How do we avoid overfitting?
- How many data examples are needed?