FOUNDATIONS OF MACHINE LEARNING

ANUSH SANKARAN

Most of the slide content are borrowed from multiple source online. Thanks to the original authors

OVERVIEW OF THE COURSE

Week	Topics
Week 1	Intro to ML Discovering ML Use Cases & ML in Business
Week 2	Python- Hands On Supervised Learning & Regression
Week 3	Neural Network - 1 Neural Network -2 (Bias, Variance) & Hands ON
Week 4	Kernel Learning & SVM Practical Advice for ML projects.
Week 5	Boosting Decision Trees, Random Forest, & xgBoost
Week 6	Unsupervised Learning Clustering & Dimensionality Reduction
Week 7	Time Series Data Analysis Imputation & Prediction Systems
Week 8	ML Use Cases from Products & Research

COURSE OUTCOMES

- Understand the fundamental concepts of different machine learning models
  - Supervised learning
  - Unsupervised learning
- Ability to formulate a business problem as machine learning task. Identify machine learning opportunities in businesses.
- Appreciate the challenging involved in data driven machine learning problems
- Ability to manage the building of tools and products that involves different aspects of machine learning

EASY LOGISTICS: GITHUB

- Github Repo: https://github.com/goodboyanush/isme-bangalore-Oct-Nov-2019
- Lectures slides, Hands-on code, Assignment solutions
- Have any doubt in my lectures or assignments?
  - Go ahead and create an issue in the repo!
  - I will try to answer them asap!
  - Everyone will be benefitted by the questions asked by one

WEEK 1:

INTRODUCTION TO MACHINE LEARNING

BUILDING THE ML MINDSET IN BUSINESS

WHAT IS MACHINE LEARNING?

- Learn a classifier: learn a mapping function

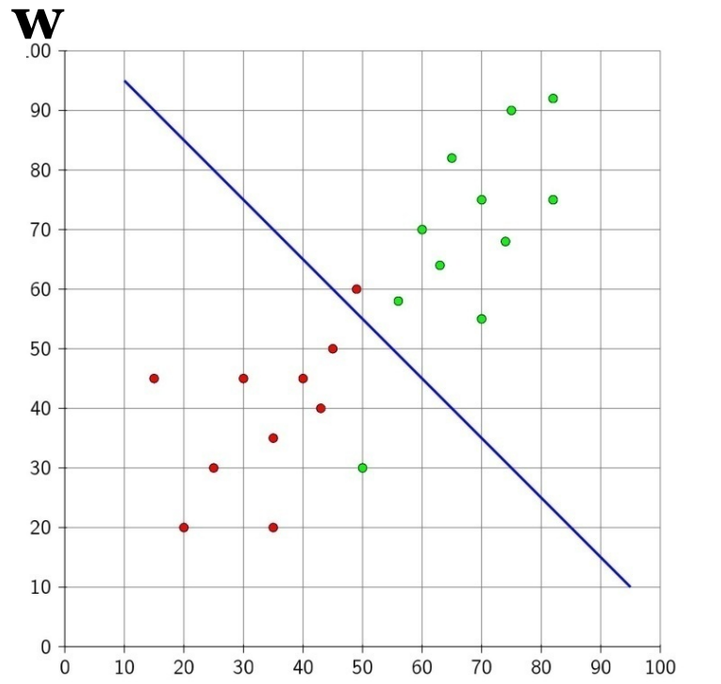
Instances/ Input Data Points

1. M: <h1, w1>
2. F: <h2, w2>
3. …

…

N. M: <hn, wn>

Labelled Features



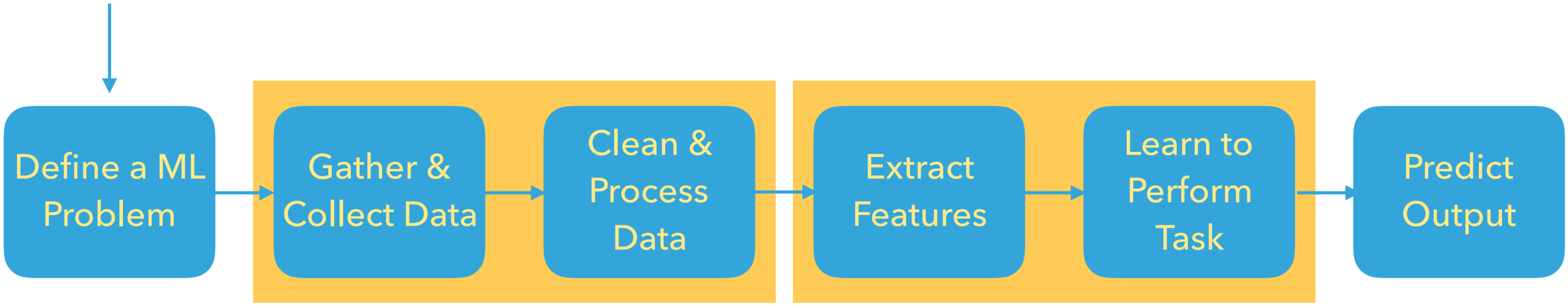
Classification

1. **Boundary**: Can be linear or non-linear boundary
2. **Method**: Can be a generative classifier or discriminative classifier

Examples: Naïve Bayes, Decision trees, Neural Network, Support Vector Machines etc.

MACHINE LEARNING PIPELINE

What are we focussing on today's lecture?



1. Articulate the problem (task)
2. Data Drive Strategy: Look for labelled data
3. Design your data for the task
4. Determine easily obtained inputs
5. Determine easily quantifiable outputs

COMMON LINGO

Input data/ Features

Output
Task Label

Instances

Loan_ID	Gender	Married	Dependents	Education	Self_Employed	ApplicantIncome	CoapplicantIncome	LoanAmount	Loan_Amount_Term	Credit_History	Property_Area	Loan_Status
LP001003	Male	Yes	1	Graduate	No	4583	1508	128	360	1	Rural	N
LP001005	Male	Yes	0	Graduate	Yes	3000	0	66	360	1	Urban	Y
LP001006	Male	Yes	0	Not Graduate	No	2583	2358	120	360	1	Urban	Y
LP001008	Male	No	0	Graduate	No	6000	0	141	360	1	Urban	Y
LP001011	Male	Yes	2	Graduate	Yes	5417	4196	267	360	1	Urban	Y
LP001013	Male	Yes	0	Not Graduate	No	2333	1516	95	360	1	Urban	Y
LP001014	Male	Yes	3+	Graduate	No	3036	2504	158	360	0	Semiurban	N
LP001018	Male	Yes	2	Graduate	No	4006	1526	168	360	1	Urban	Y
LP001020	Male	Yes	1	Graduate	No	12841	10968	349	360	1	Semiurban	N
LP001024	Male	Yes	2	Graduate	No	3200	700	70	360	1	Urban	Y
LP001028	Male	Yes	2	Graduate	No	3073	8106	200	360	1	Urban	Y
LP001029	Male	No	0	Graduate	No	1853	2840	114	360	1	Rural	N
LP001030	Male	Yes	2	Graduate	No	1299	1086	17	120	1	Urban	Y
LP001032	Male	No	0	Graduate	No	4950	0	125	360	1	Urban	Y
LP001036	Female	No	0	Graduate	No	3510	0	76	360	0	Urban	N

Seen data / Training data

Unseen data/ Test data

LP001038	Male	Yes	0	Not Graduate	No	4887	0	133	360	1	Rural
LP001043	Male	Yes	0	Not Graduate	No	7660	0	104	360	0	Urban
LP001046	Male	Yes	1	Graduate	No	5955	5625	315	360	1	Urban
LP001047	Male	Yes	0	Not Graduate	No	2600	1911	116	360	0	Semiurban

Machine Learning Glossary: https://developers.google.com/machine-learning/glossary

THE ML MINDSET

"Machine Learning changes the way you think about a problem.

The focus shifts from a mathematical science to a natural science, running experiments and using statistics, not logic, to analyse its results."

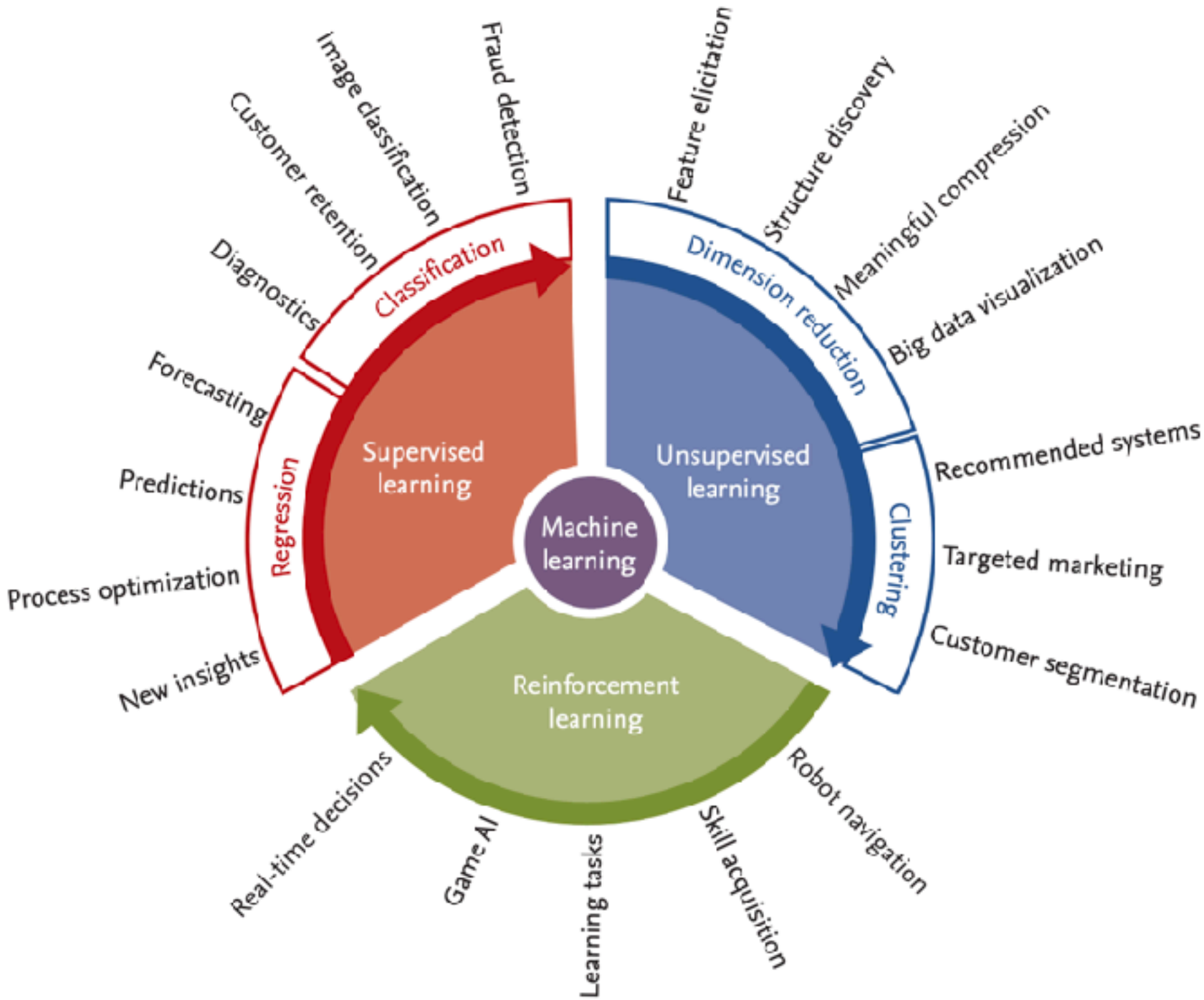
- **Peter Norvig**

THE ML MINDSET

Step	Example
1. Set the research goal.	I want to predict how heavy traffic will be on a given day.
2. Make a hypothesis.	I think the weather forecast is an informative signal.
3. Collect the data.	Collect historical traffic data and weather on each day.
4. Test your hypothesis.	Train a model using this data.
5. Analyze your results.	Is this model better than existing systems?
6. Reach a conclusion.	I should (not) use this model to make predictions, because of X, Y, and Z.
7. Refine hypothesis and repeat.	Time of year could be a helpful signal.

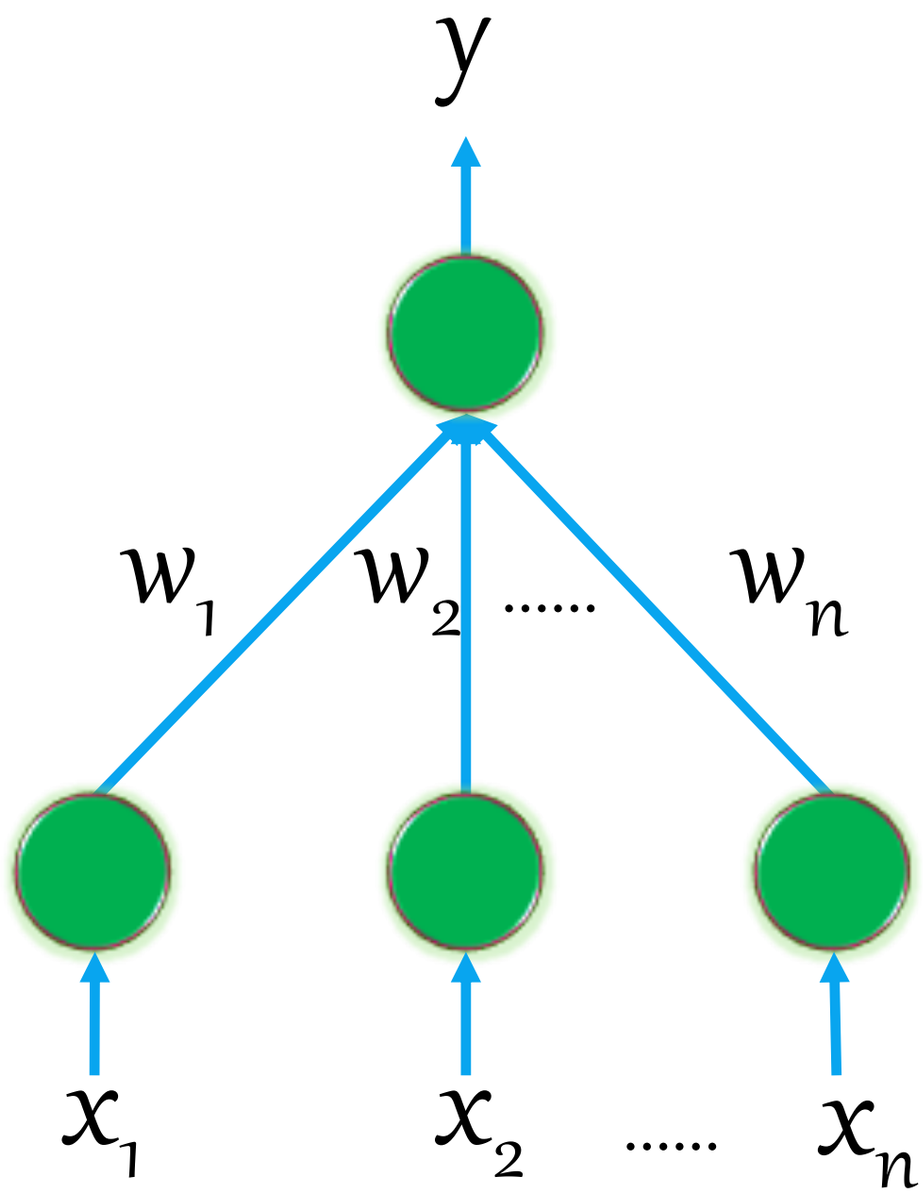
Get Comfortable with Some Uncertainty !

DIFFERENT TYPES OF ML ALGORITHMS



LINEAR REGRESSION – ACTIVATION FUNCTION

$$y = f\left(\Sigma_i w_i x_i\right)$$



$$\text{Error} = (y - y')^2$$

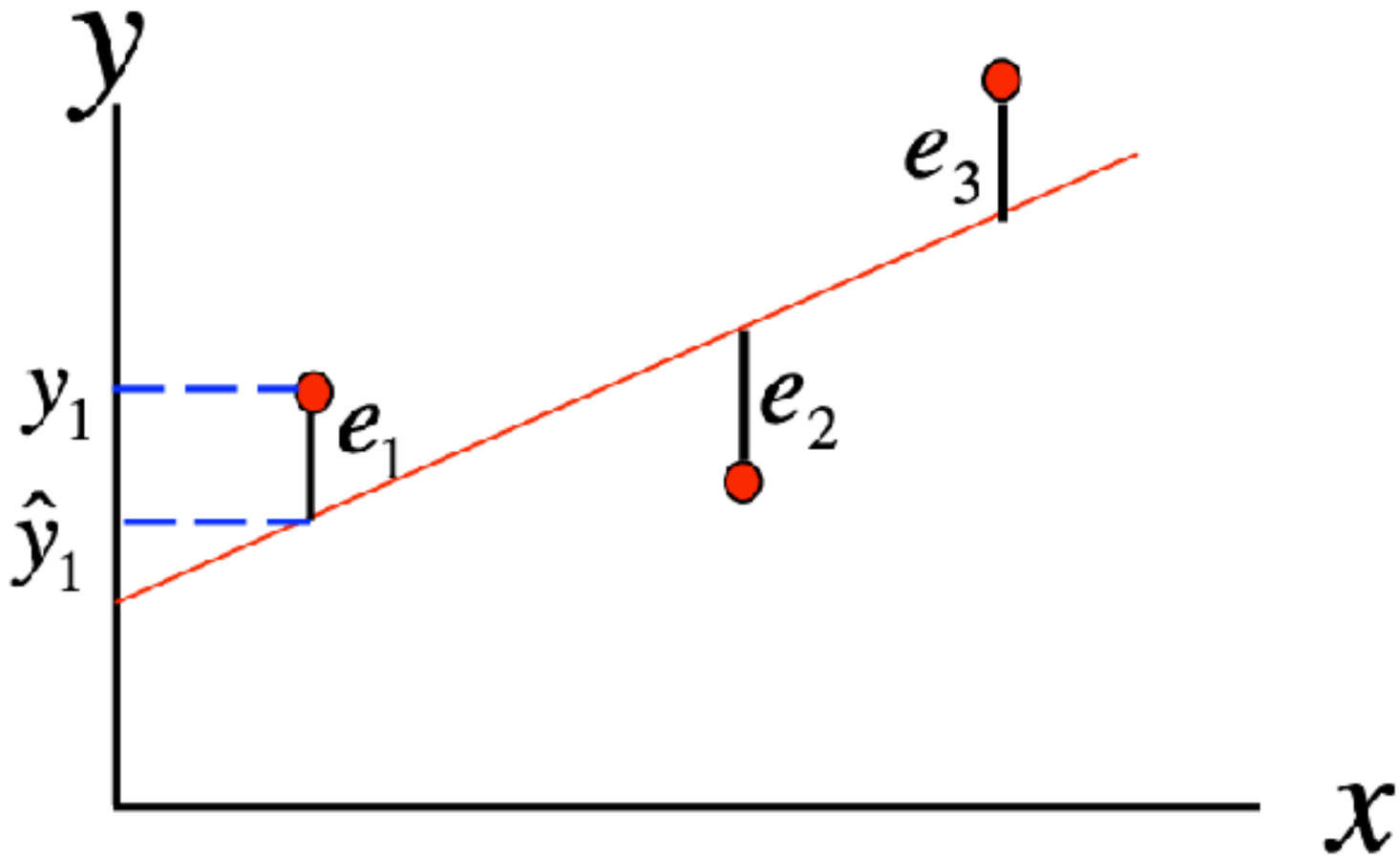
Residual Sum of Squares (RSS)

Update **w** in such a way that the error is minimized !

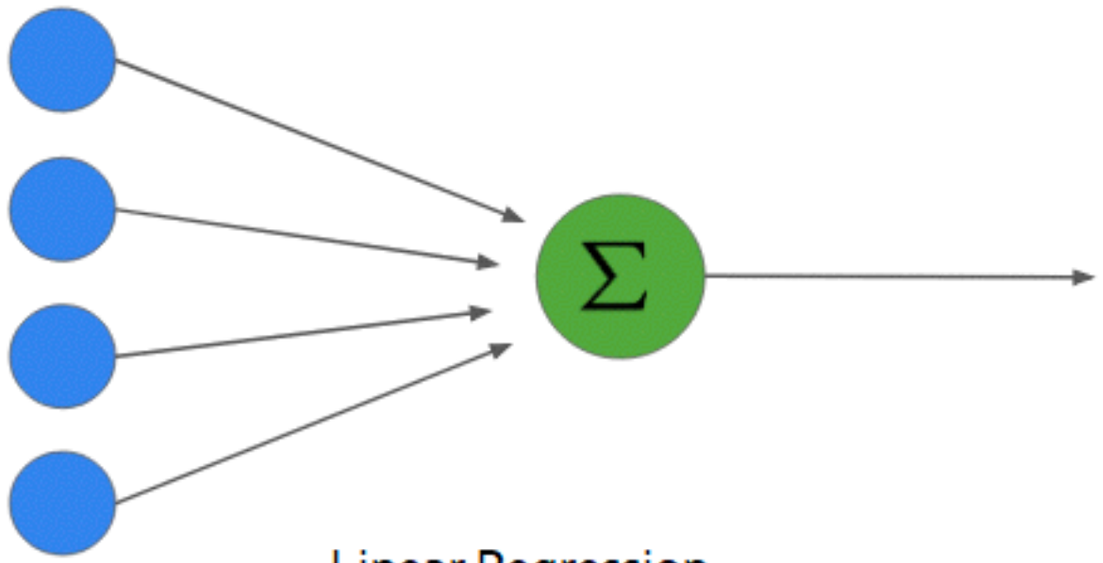
RESIDUAL SUM OF SQUARES

$$e_1 = y_1 - \hat{y}_1$$

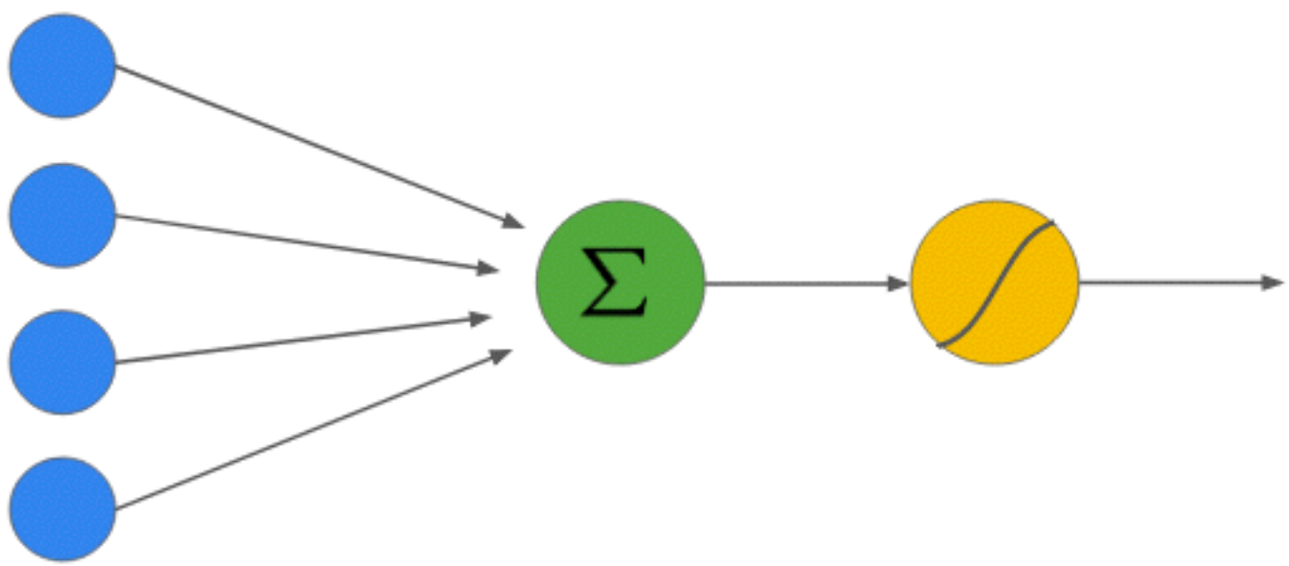
$$e_2 = y_2 - \hat{y}_2$$



LINEAR VS. LOGISTIC REGRESSION

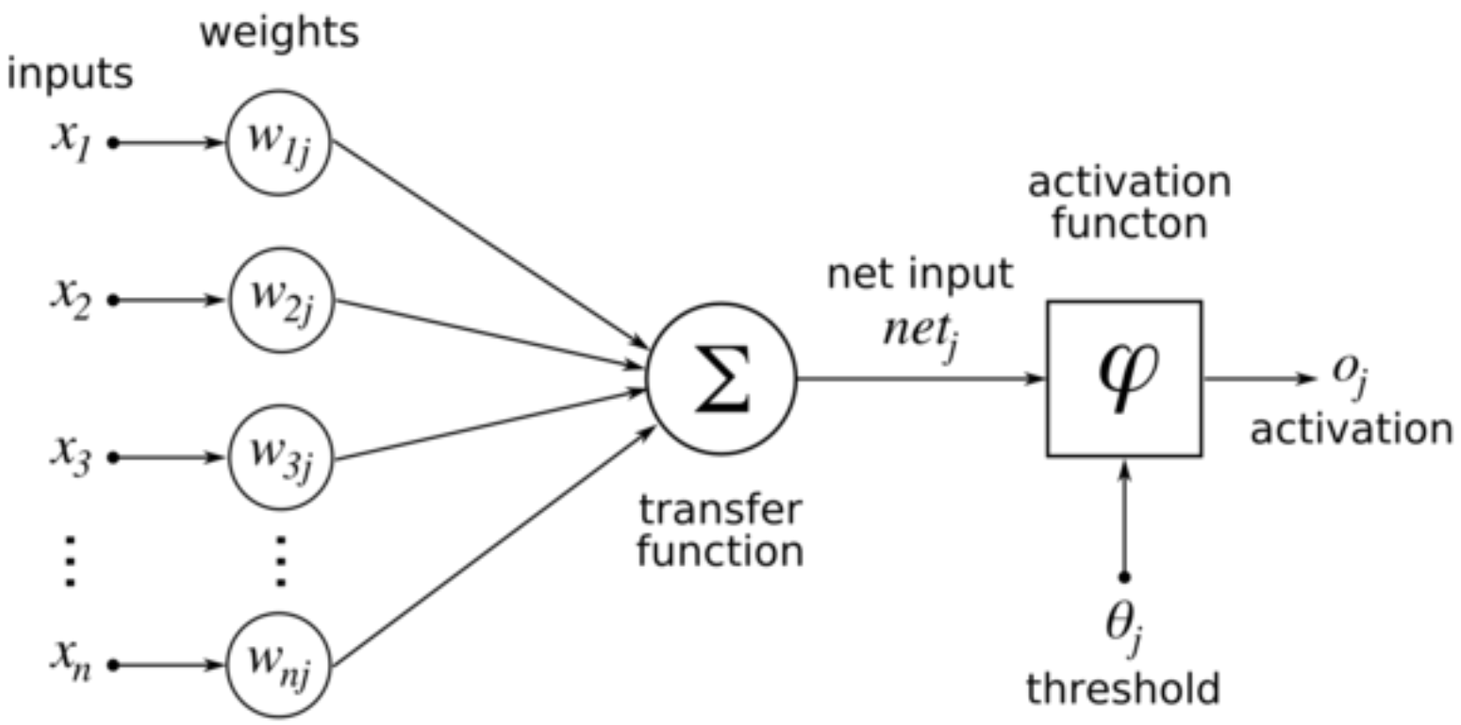


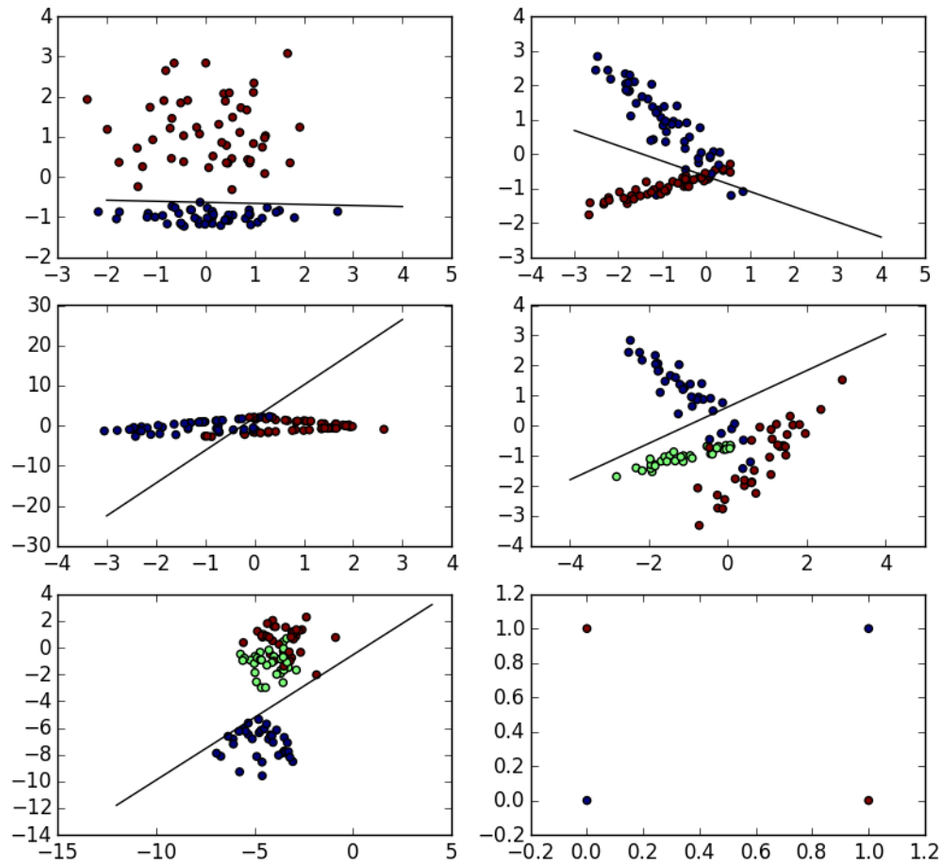
Linear Regression



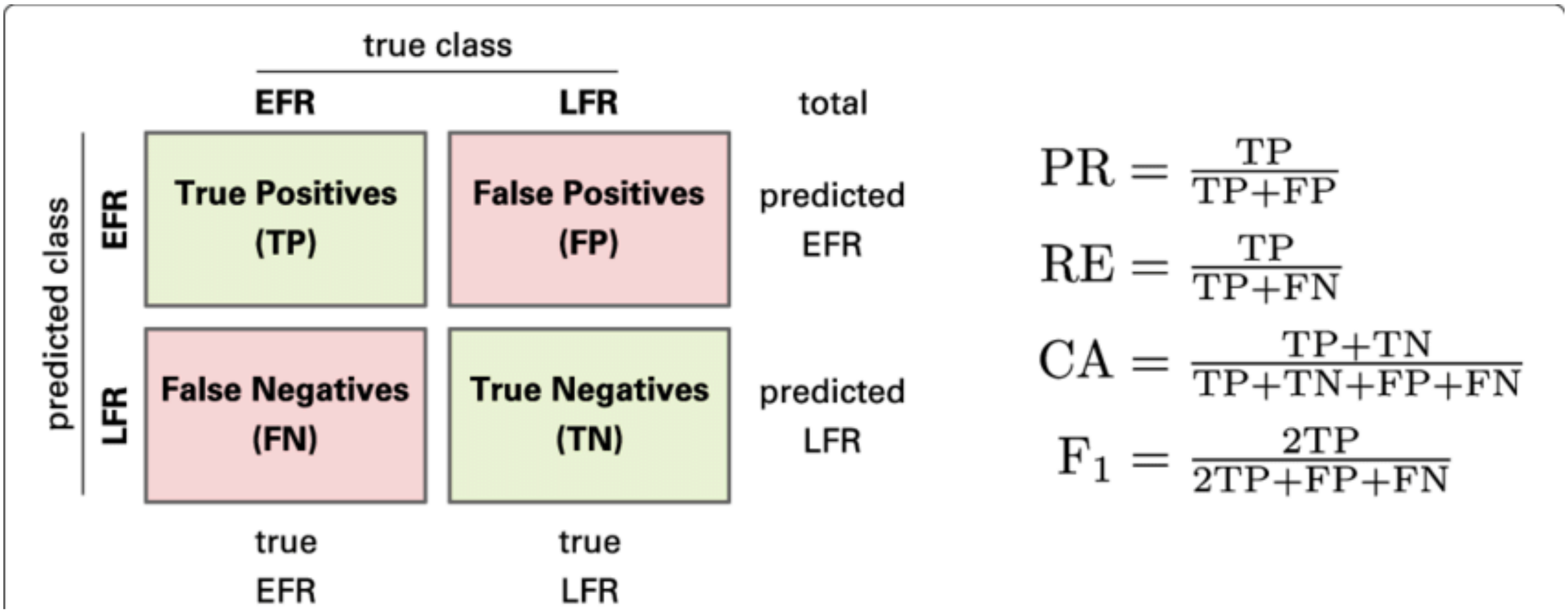
Logistic Regression

LOGISTIC REGRESSION





METRICS FOR MEASUREMENT



$$\mathrm{PR} = \frac{\mathrm{TP}}{\mathrm{TP+FP}}$$

$$\mathrm{RE} = \frac{\mathrm{TP}}{\mathrm{TP+FN}}$$

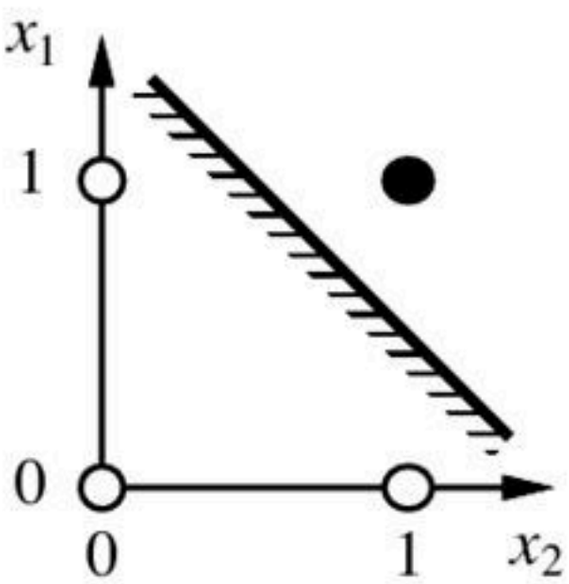
$$\mathrm{CA} = \frac{\mathrm{TP+TN}}{\mathrm{TP+TN+FP+FN}}$$

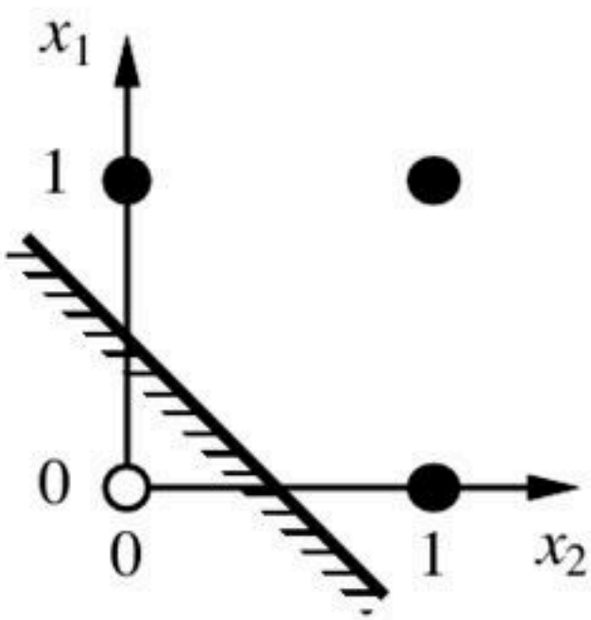
$$\mathrm{F}_1 = \frac{2\mathrm{TP}}{2\mathrm{TP+FP+FN}}$$

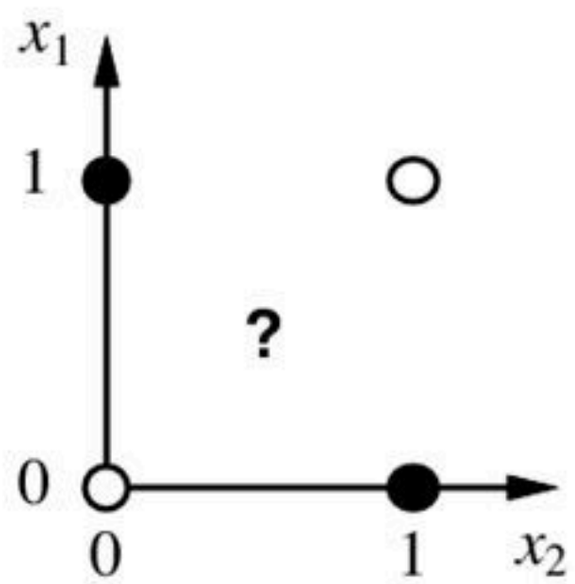
THANK YOU – NEXT WEEK

Week	Topics
Week 1	Intro to ML Discovering ML Use Cases & ML in Business
Week 2	Python- Hands On Supervised Learning & Regression
Week 3	Neural Network - 1 Neural Network -2 (Bias, Variance) & Hands ON
Week 4	Kernel Learning & SVM Practical Advice for ML projects.
Week 5	Boosting Decision Trees, Random Forest, & xgBoost
Week 6	Unsupervised Learning Clustering & Dimensionality Reduction
Week 7	Time Series Data Analysis Imputation & Prediction Systems
Week 8	ML Use Cases from Products & Research

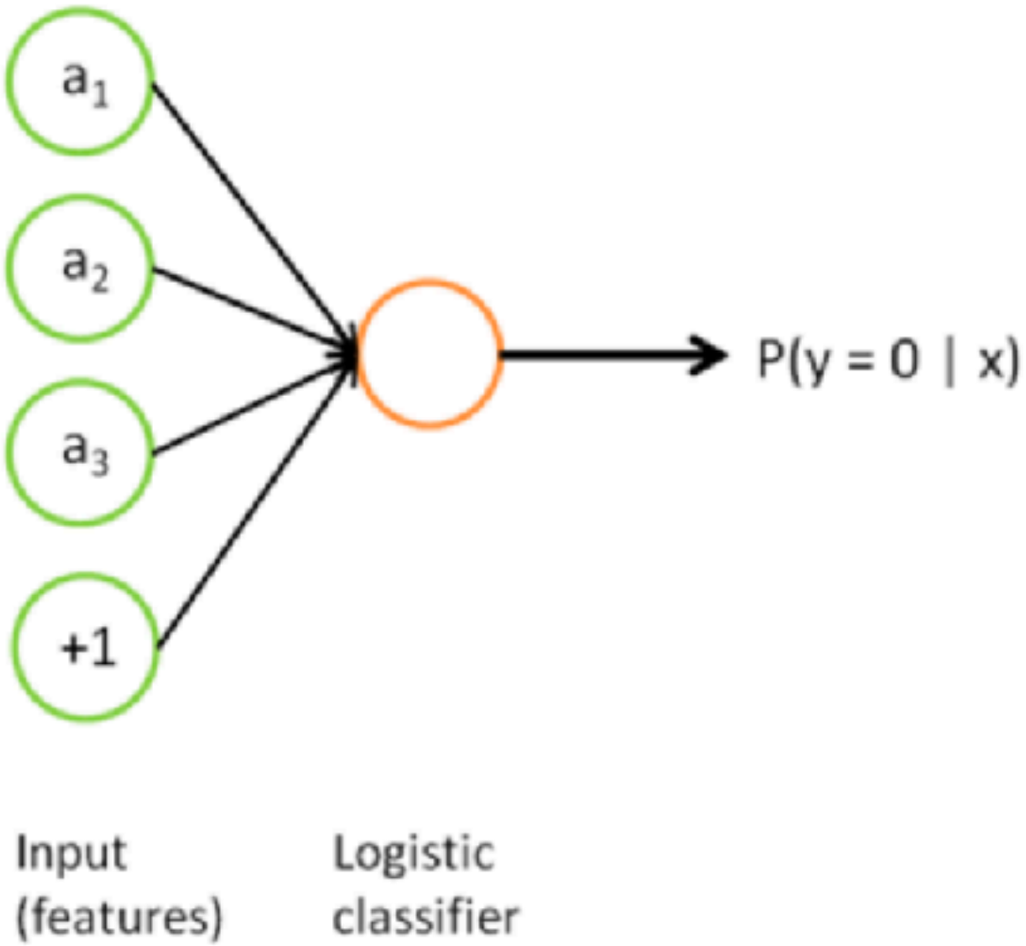
LOGISTIC REGRESSION – LINEAR DECISION BOUNDARY



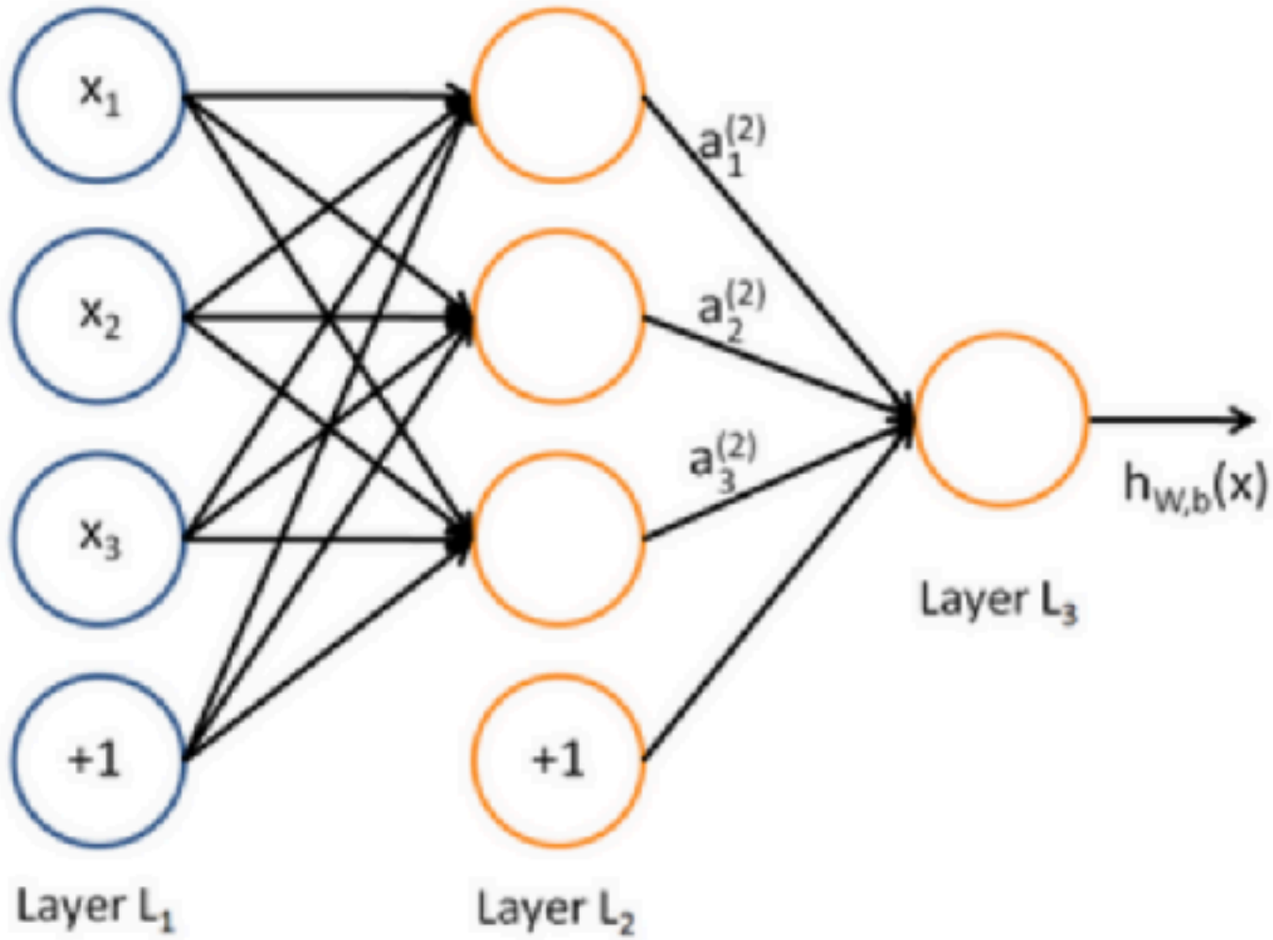




NEURAL NETWORKS

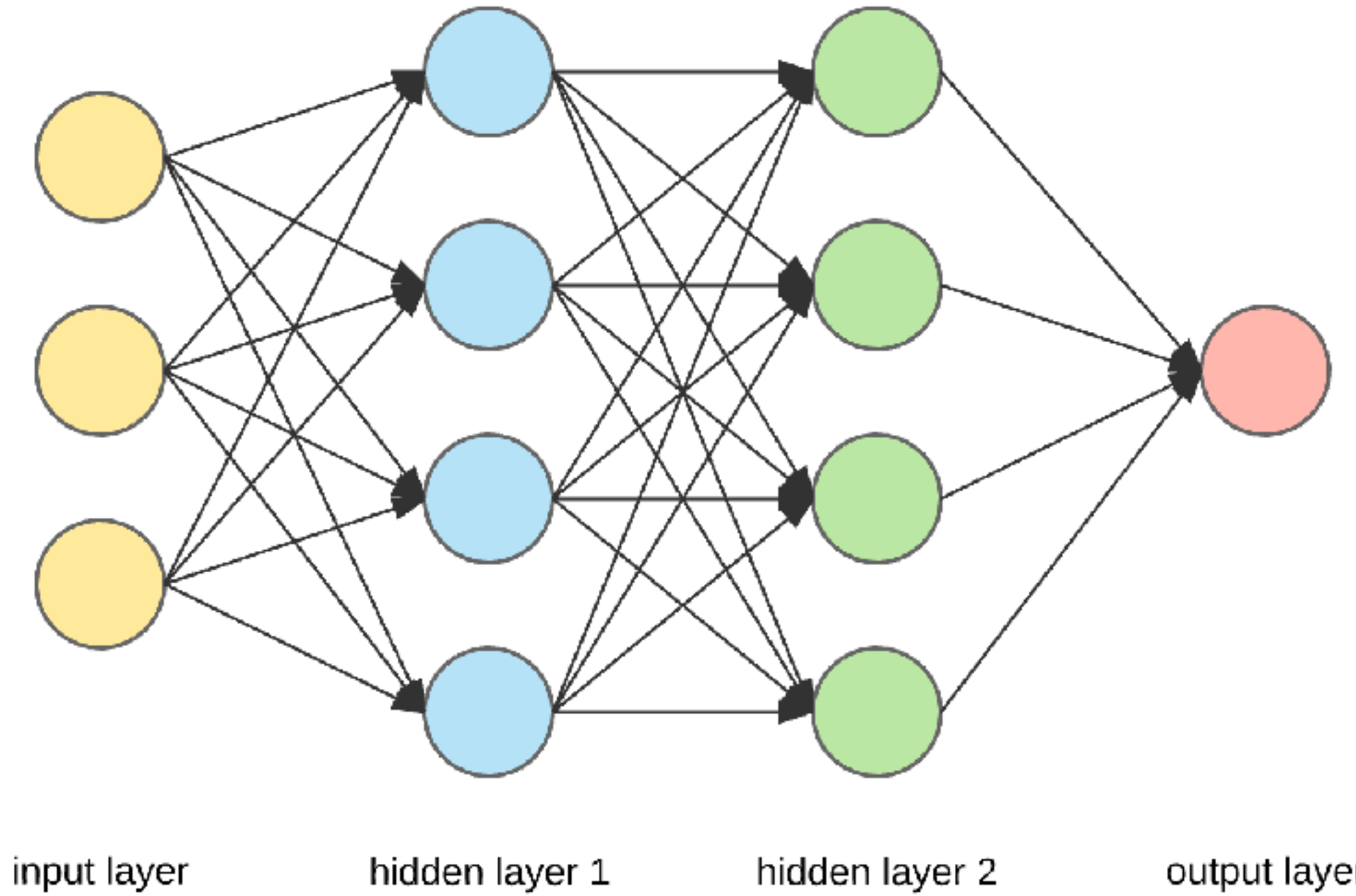


Logistic Regression

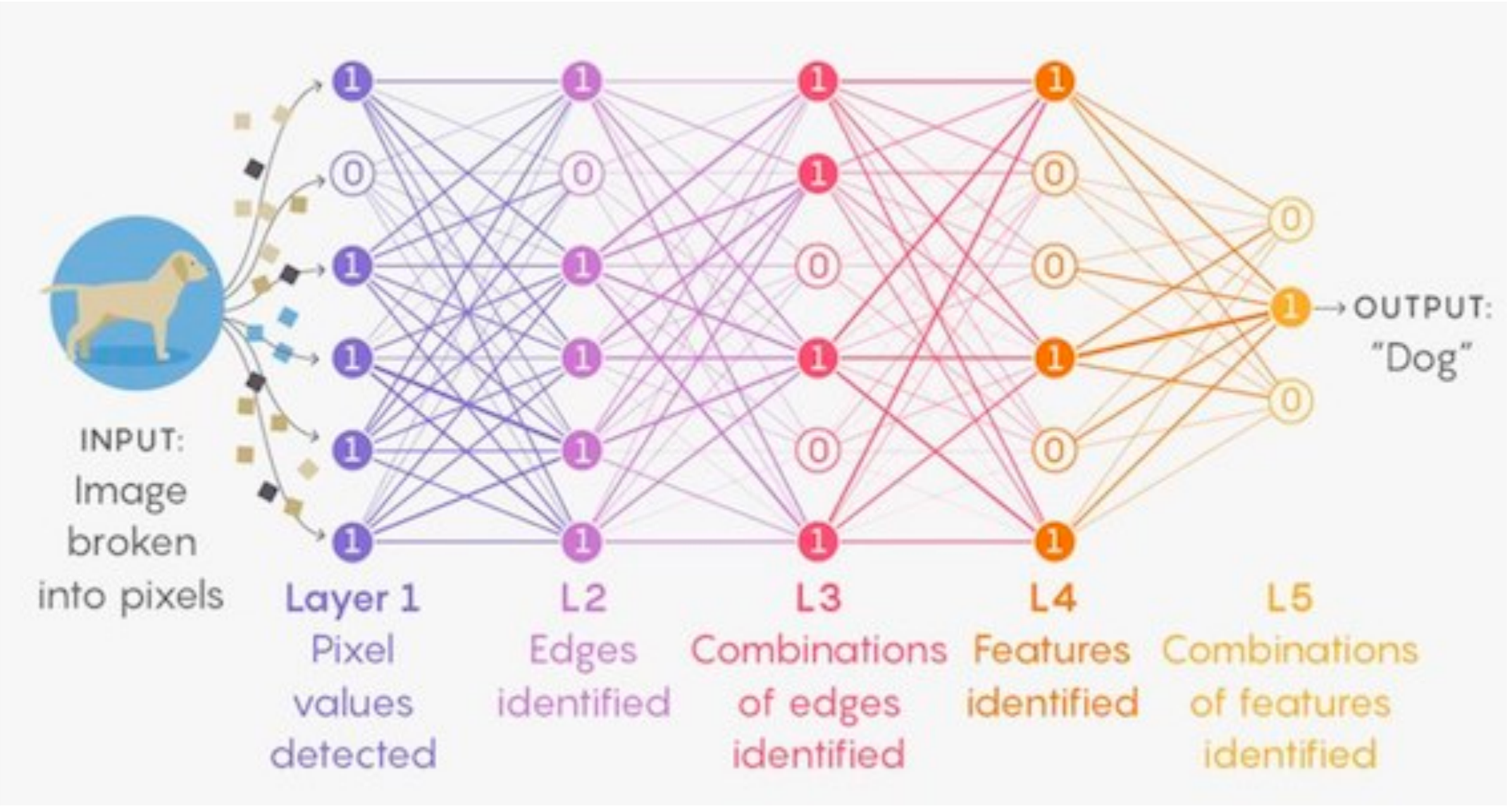


Neural Network

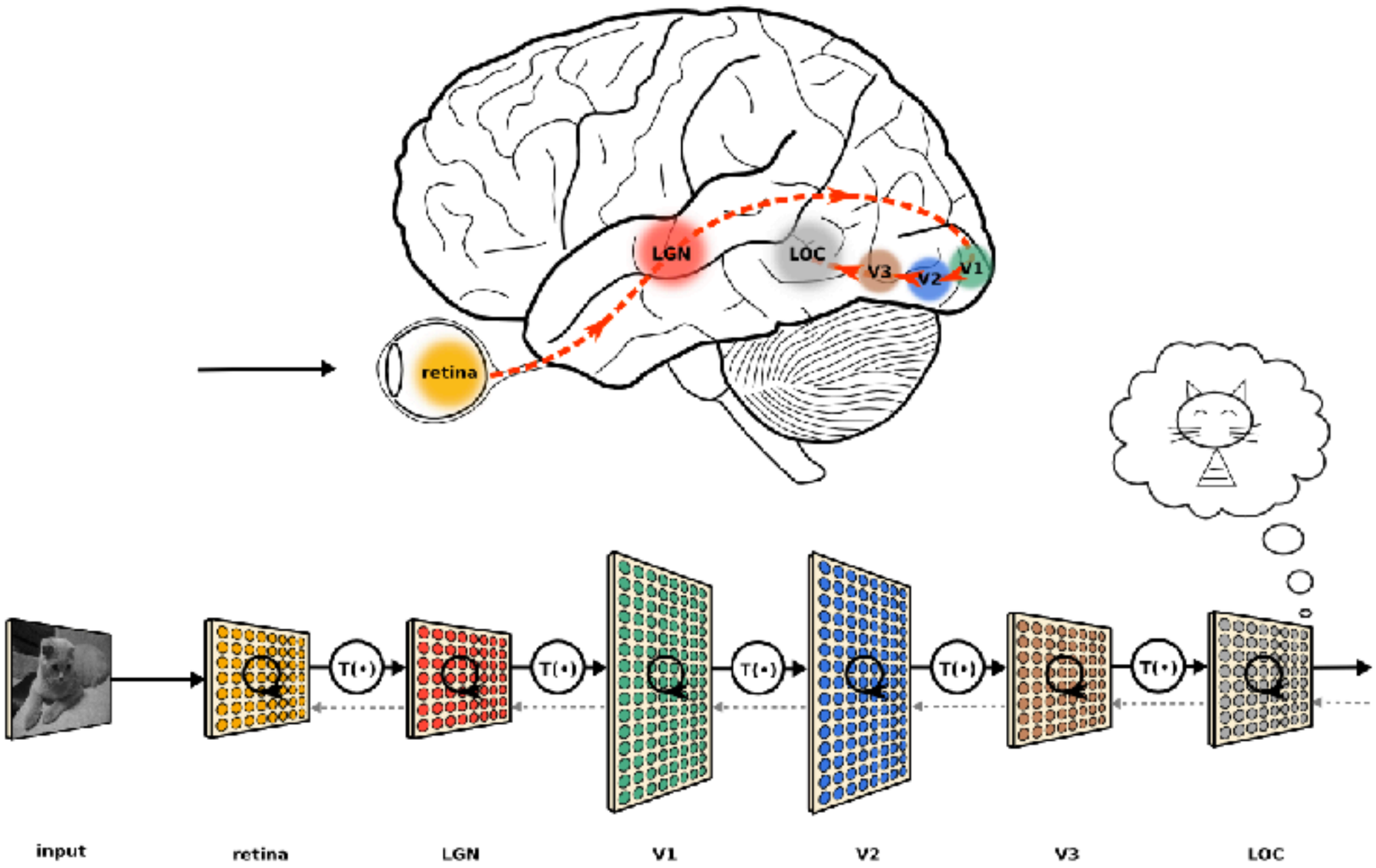
NEURAL NETWORKS



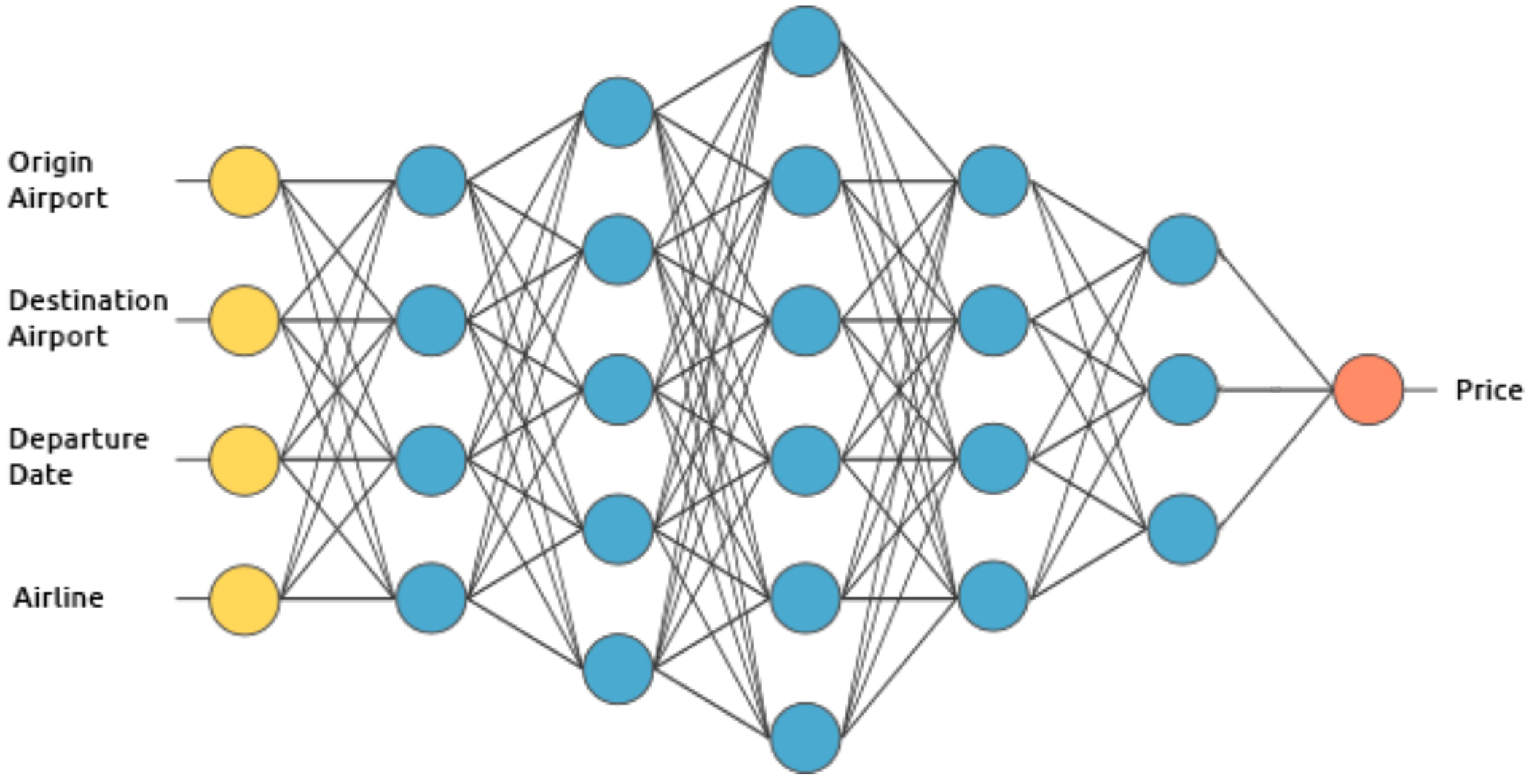
WHAT IS IT LEARNING?



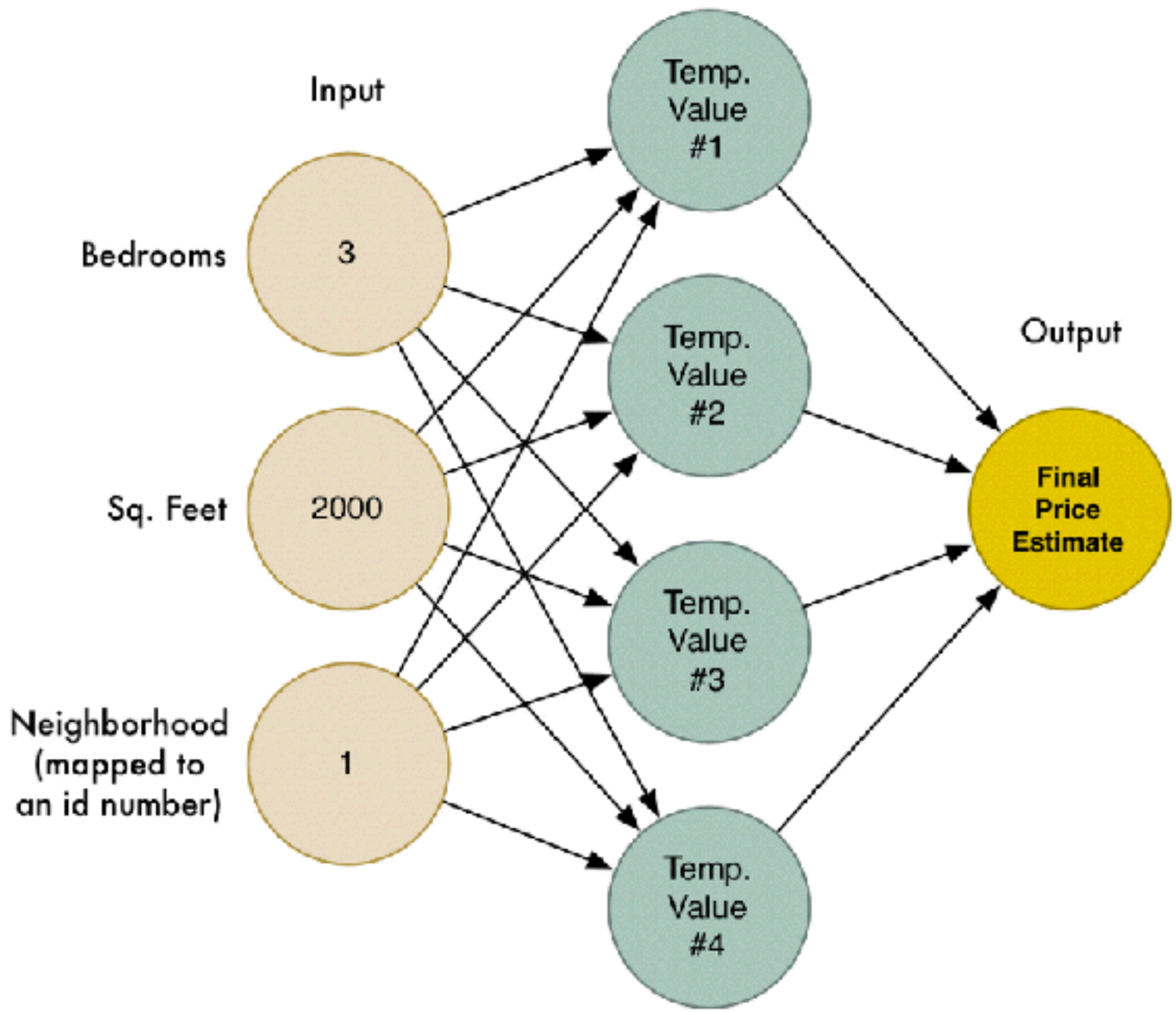
NEURAL NETWORK AND HUMAN BRAIN



EXAMPLE 1: FLIGHT PRICE PREDICTION



EXAMPLE 2: HOUSE RENT PREDICTION



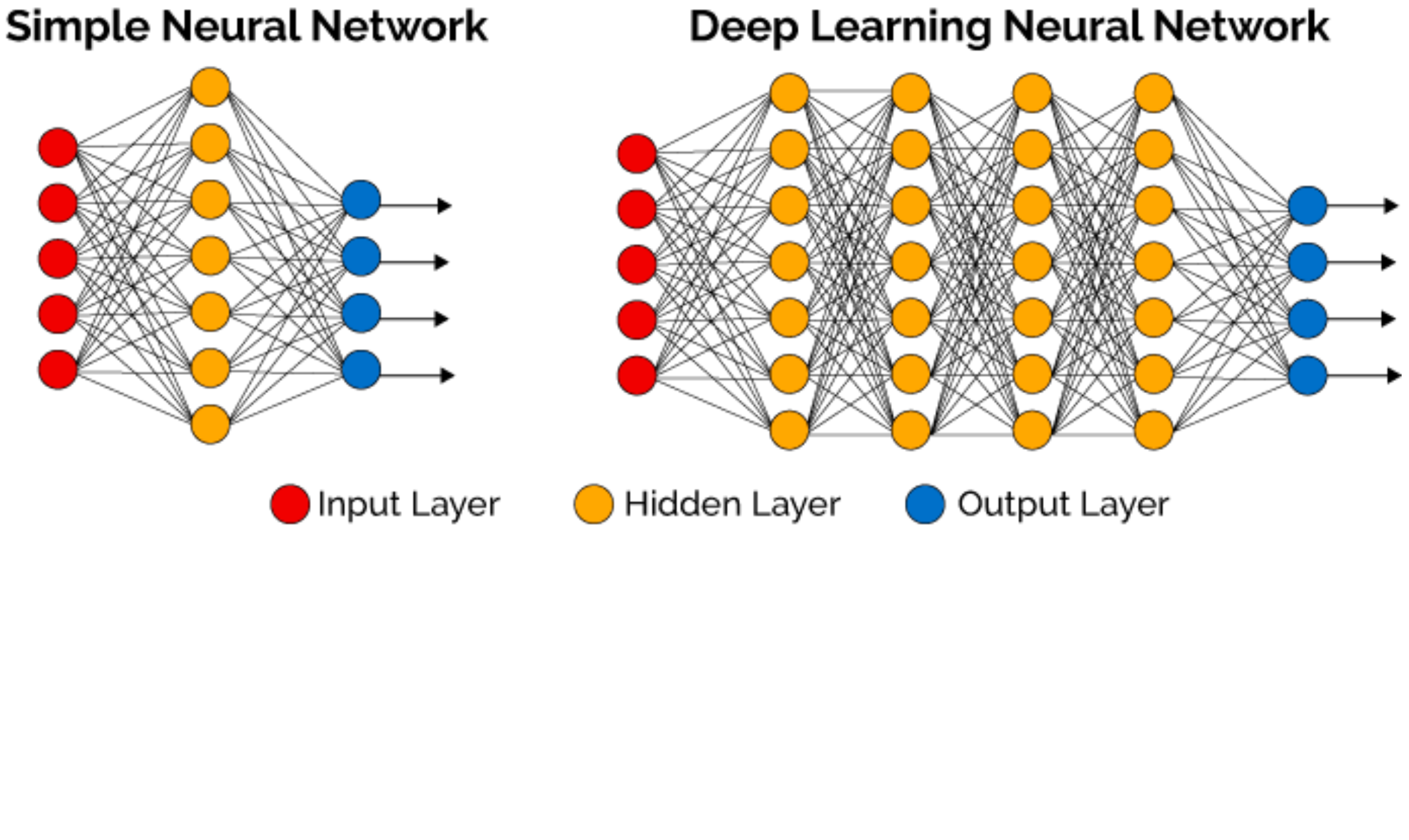
VARIANTS OF NEURAL NETWORK

- Input data is a vector
  - Stacked Autoencoder
  - Restricted Boltzmann Machines
  - Deep Belief Networks

- Input data is multi-dimensional
  - Convolutional Neural Network
  - Example: Image analysis, Object recognition

- Input data is time-series
  - Recurrent Neural Network
  - Example: Language Modelling, Speech recognition

DEEP LEARNING



LETS DO CODING

TASKS FOR THIS WEEK

- Write a blog post on:
  - What is the difference between regression and neural network

THANK YOU – NEXT WEEK

Week	Topics
Week 1	Intro to ML Discovering ML Use Cases & ML in Business
Week 2	Python- Hands On Supervised Learning & Regression
Week 3	Neural Network - 1 Neural Network -2 (Bias, Variance) & Hands ON
Week 4	Kernel Learning & SVM Practical Advice for ML projects.
Week 5	Boosting Decision Trees, Random Forest, & xgBoost
Week 6	Unsupervised Learning Clustering & Dimensionality Reduction
Week 7	Time Series Data Analysis Imputation & Prediction Systems
Week 8	ML Use Cases from Products & Research