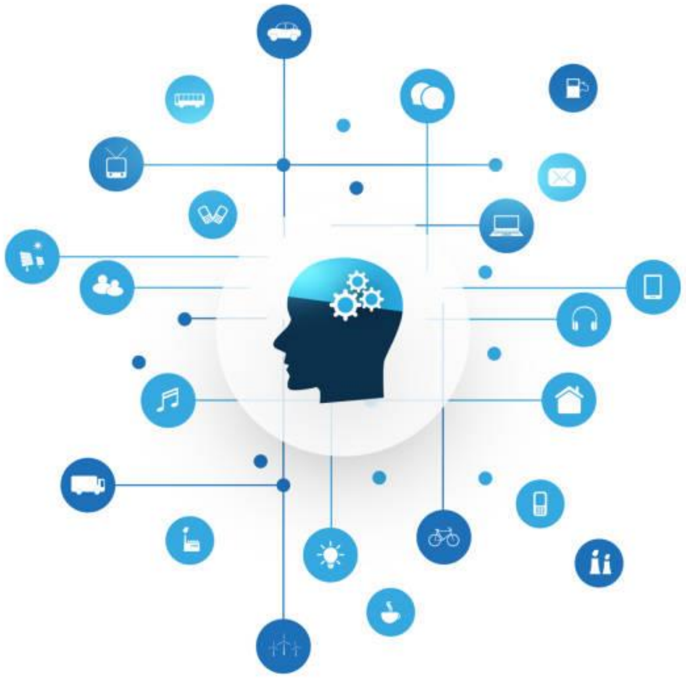
Introduction to Machine Learning for Beginners



AI SCIENCES

Some Applications

Why Now?

➔ Availability of huge amount of data

Why Now?

- ➔ Availability of huge amount of data
- ➔Powerful machines that can handle computations quickly

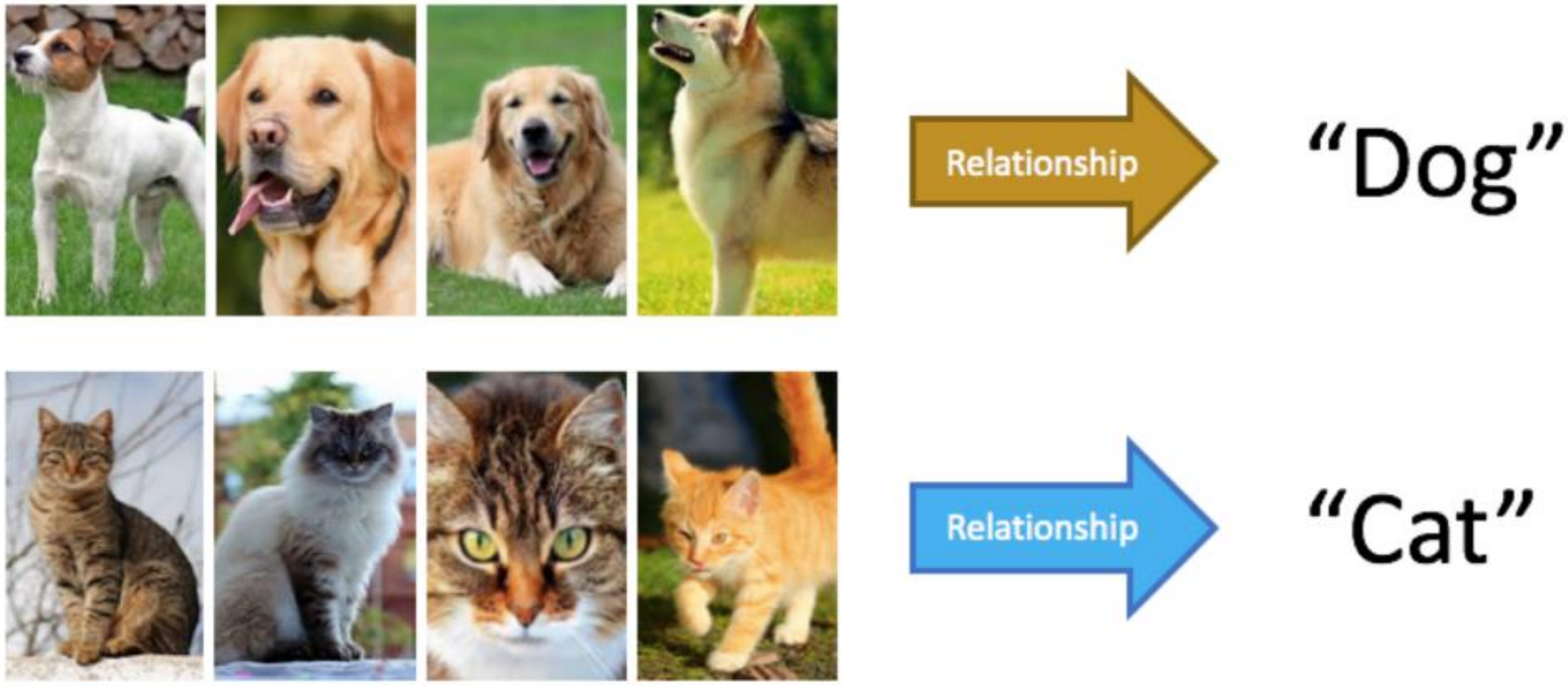
Why Now?

- ➔ Availability of huge amount of data
- ➔ Powerful machines that can handle computations quickly
- ➔ Advancement of technology and knowledge among people

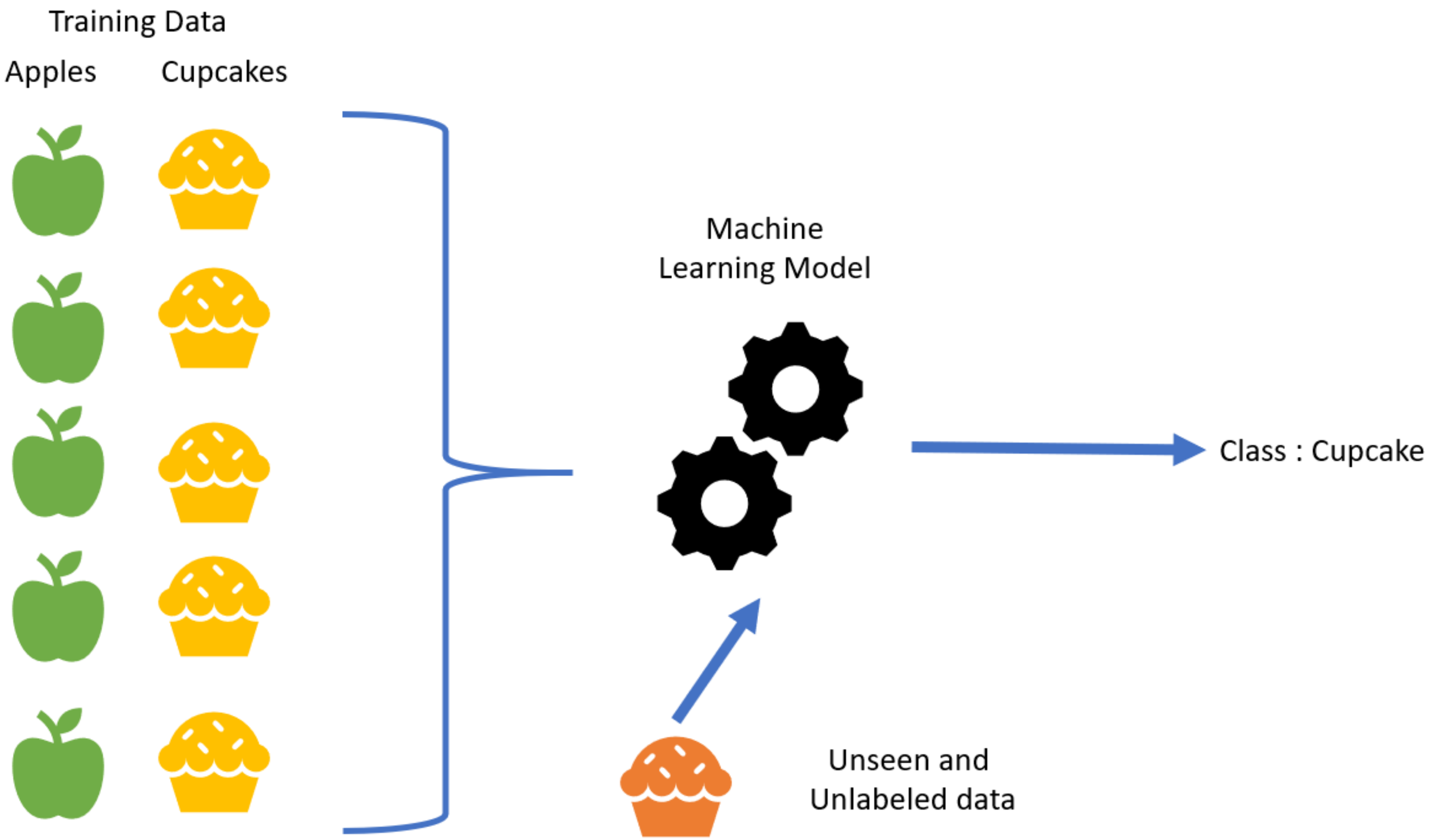
Learning from Data

- Supervised Learning
- Unsupervised Learning
- Reinforcement Learning

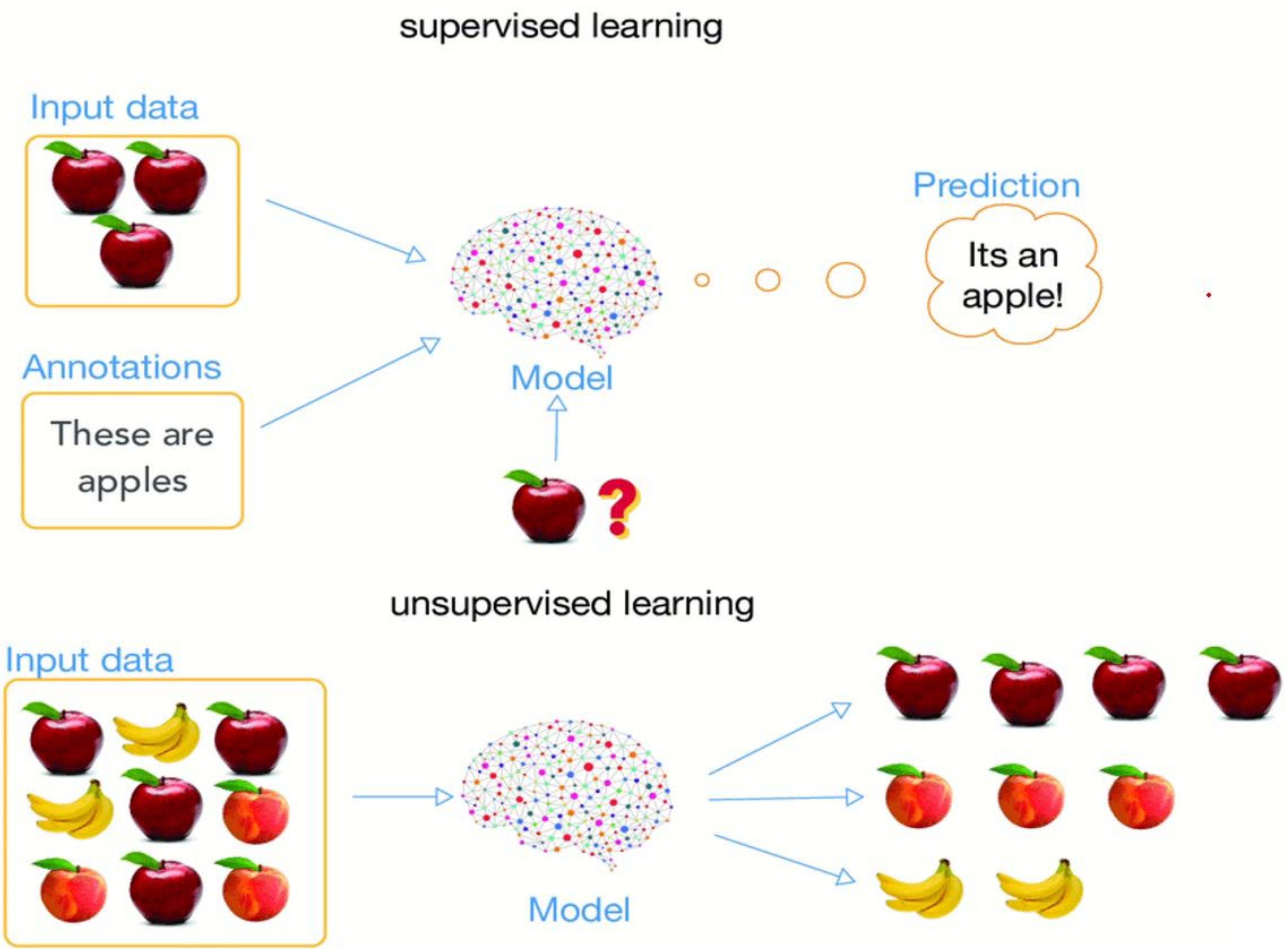
Supervised Learning



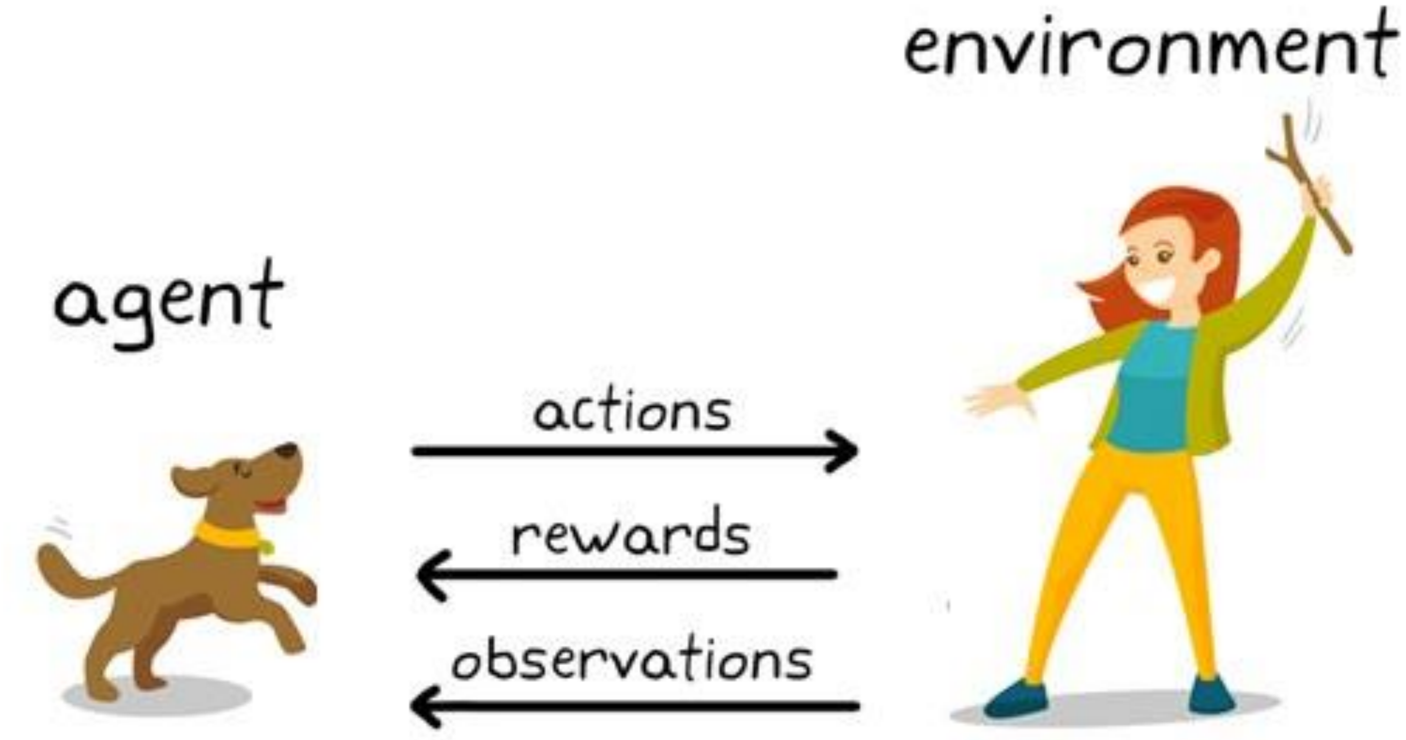
Supervised Learning



Unsupervised Learning



Reinforcement Learning



Features

Regression

Classification

Clustering

Data Encoding and Preprocessing

➔ Images

Data Encoding and Preprocessing

- ➔ Images
- ➔ Videos

Data Encoding and Preprocessing

- ➔ Images
- ➔ Videos
- ➔ Audio

Data Encoding and Preprocessing

- ➔ Images
- ➔ Videos
- ➔ Audio
- ➔ Text
- ➔ One Hot encoding

Data Encoding and Preprocessing

- ➔ Images
- ➔ Videos
- ➔ Audio
- ➔ Text
- ➔ One Hot encoding
- ➔ Data standardization

Model

➔ Model is a function in the feature space

Model

- Model is a function in the feature space
- Dimensions
- Parameters
- Hyperparameters

Error/Cost and Optimization

Result of training: values of parameters.

x_1	x_2	y
2	10	-1
3	-4	1
-3	100	-1
10	20	1

$\hat{y}_i - y_i == 0$

$$\hat{y}_i = f(x_{1i}, x_{2i}) = a x_{1i} + b x_{2i} + c$$

$$\boxed{\hat{y}_i = f(x_{1i}, x_{2i})}$$

$$\boxed{\hat{y}_i == y_i}$$

y_Pred ,

$$\boxed{\hat{y}_i, y_i}$$

Error/Cost and Optimization

$$\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2.$$

Linear Regression

x_i	y_i
2	5
-3	-2
7	5.9
.	.
.	.
.	.

$$\boxed{\hat{y}_i = a x_i + b}$$

$$a x_i + b = y_i$$

$$\begin{bmatrix} x_i & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} y_i \end{bmatrix}$$

Linear Regression

$$a x_i + b = y_i$$

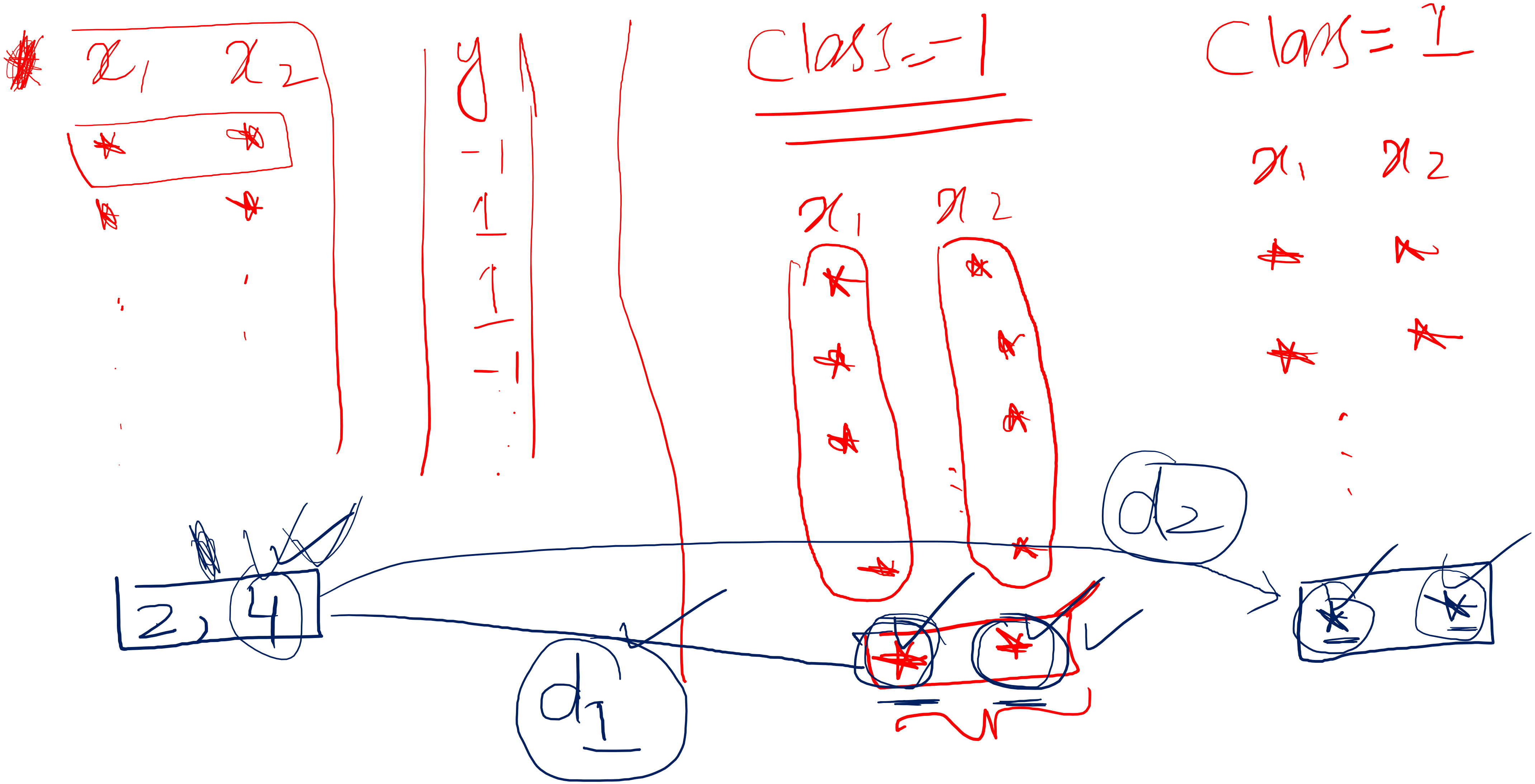
$$\begin{bmatrix} x_i & 1 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = y_i$$

$$\underset{n\times 2}{X} \underset{2\times 1}{a} = \underset{n\times 1}{y}$$

$$\boxed{\begin{bmatrix} x_1 & 1 \\ x_2 & 1 \\ x_3 & 1 \\ & \\ x_n & 1 \end{bmatrix}} \boxed{\begin{bmatrix} a \\ b \end{bmatrix}} = \boxed{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}}$$

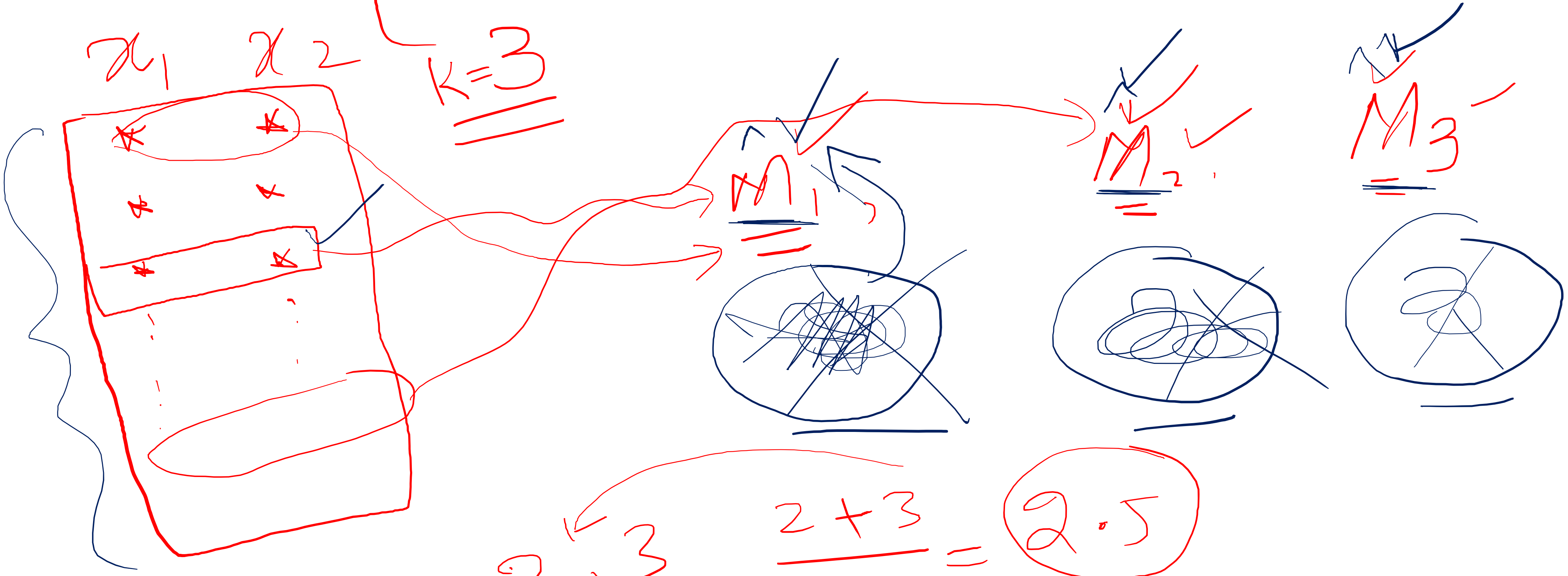
$$\boxed{X a = y}$$

Minimum-to-mean distance Classifier



K-means Clustering

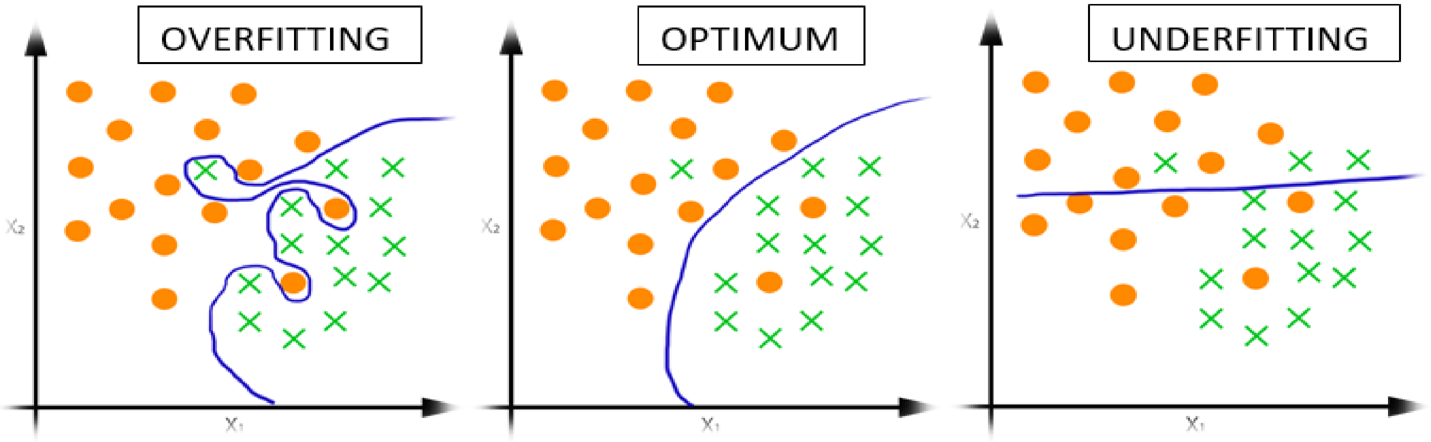
NP-hard

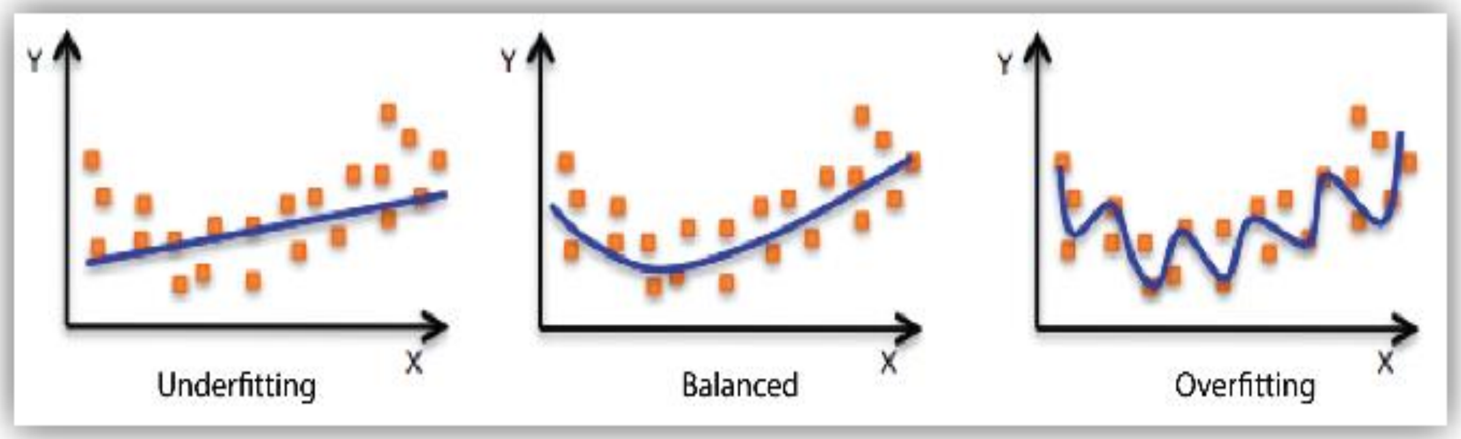


$2, 3$

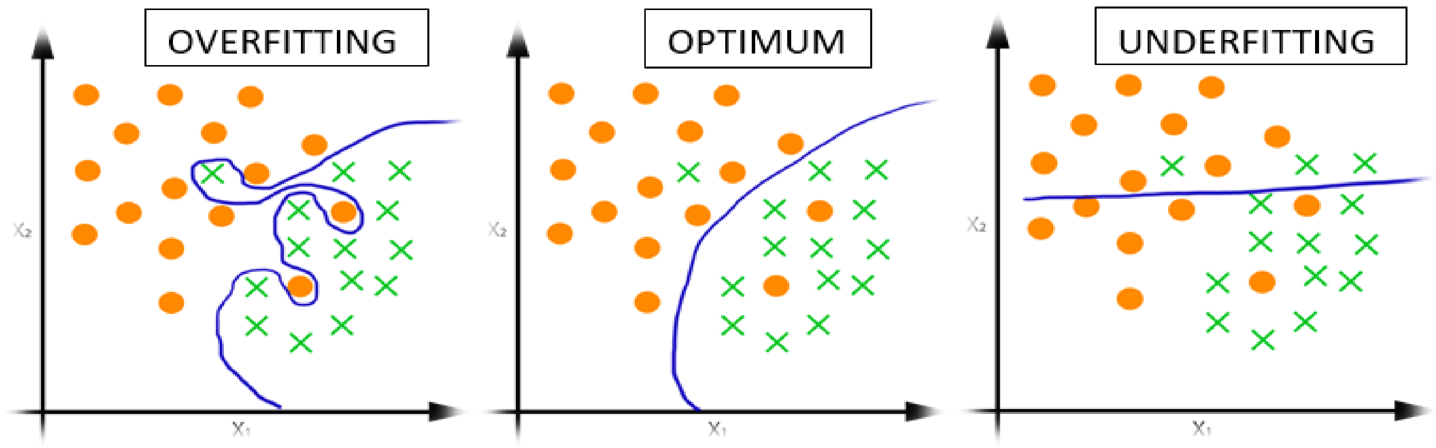
$$\frac{2+3}{2} = 2.5$$

Overfitting

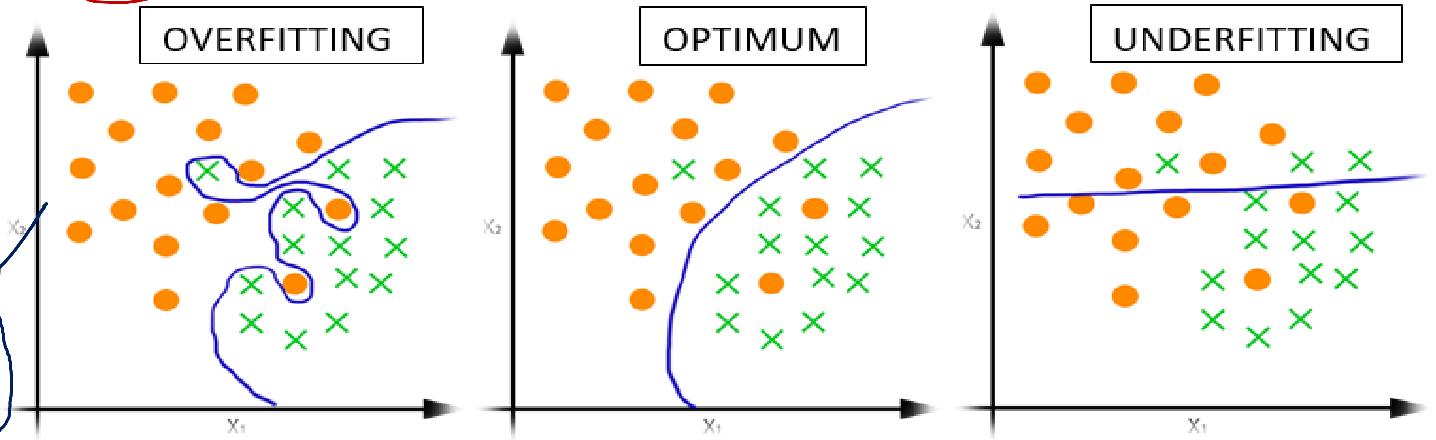




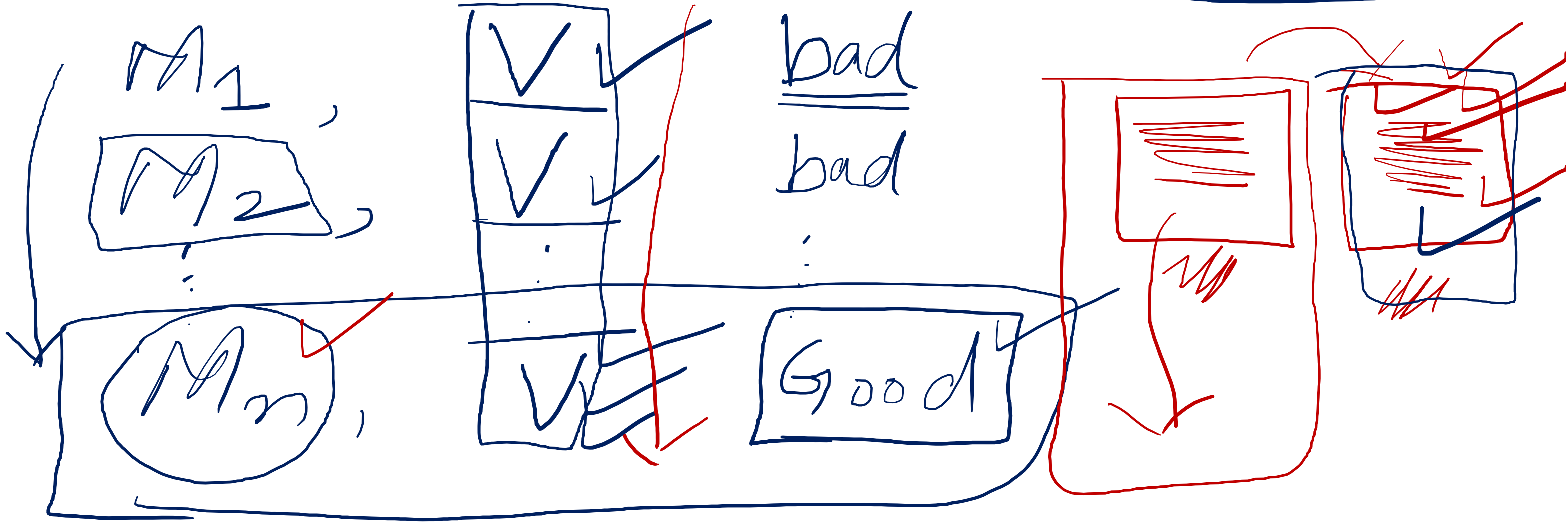
Generalization and Validation Set



Data Snooping and the Test set

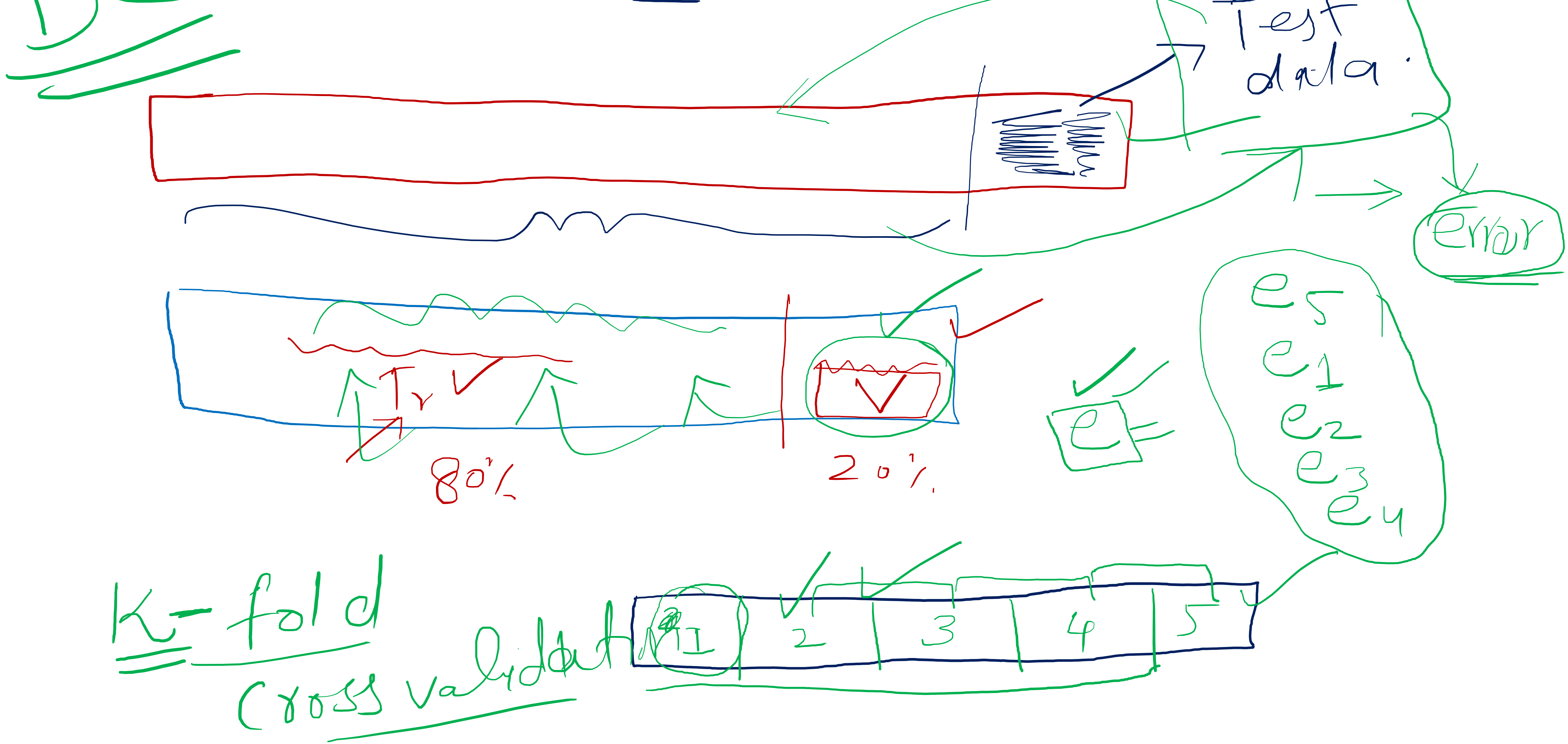


If you torture the data long enough, it will confess.



Cross Validation

BE



Performance measures

Accuracy: Error (MSE)

$(\hat{y}-y)^2$

$$Acc = \frac{\cancel{num}\ \#\text{ of true predictions}}{\#\text{ of Samples}} \quad \frac{80}{100}$$

80%

Patient with ~~cancer~~ → NO

Patient with No C → Yes

Performance measures

Confusion Matrix:

	c_1	c_2	c_3
c_1	$\boxed{a}$	b	c
c_2	d	e	f
c_3	g	h	k

Precision

$$\cancel{\frac{TP}{TP+F}}$$

$$\text{Recall} = \frac{a}{a+b}$$

	P_1	P_2
C_1	a	b
C_2	c	d

$$\frac{a}{a+c}$$

Probability Distributions and Curse of Dimensionality

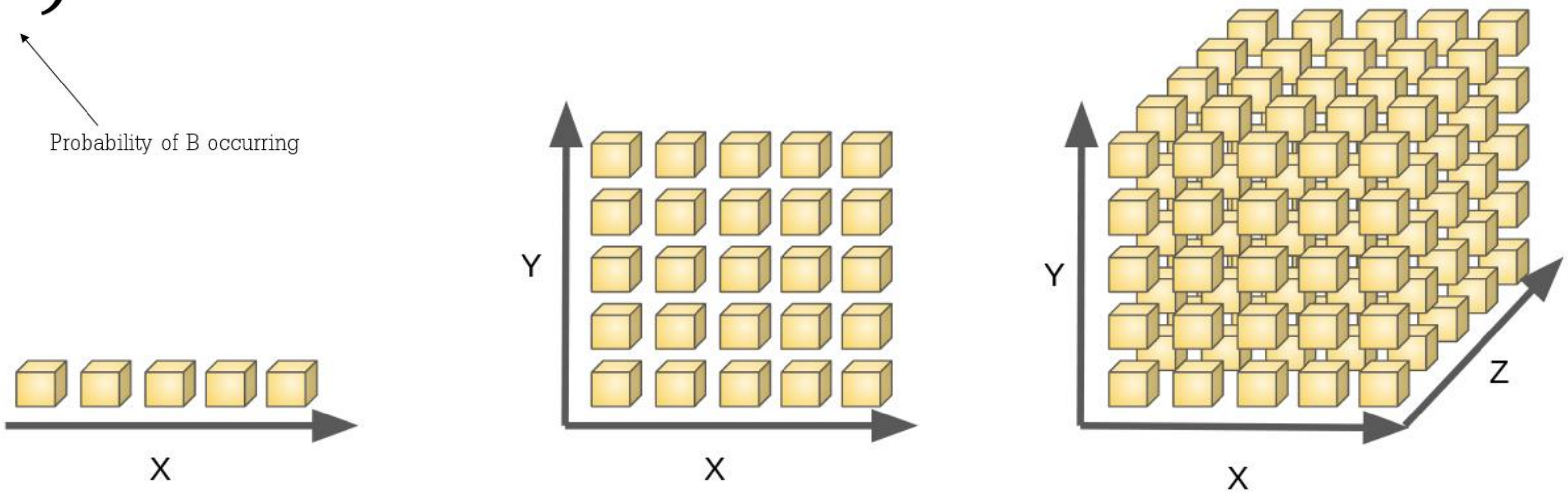
Probability of B occurring given evidence A has already occurred

Probability of A occurring

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Probability of A occurring given evidence B has already occurred

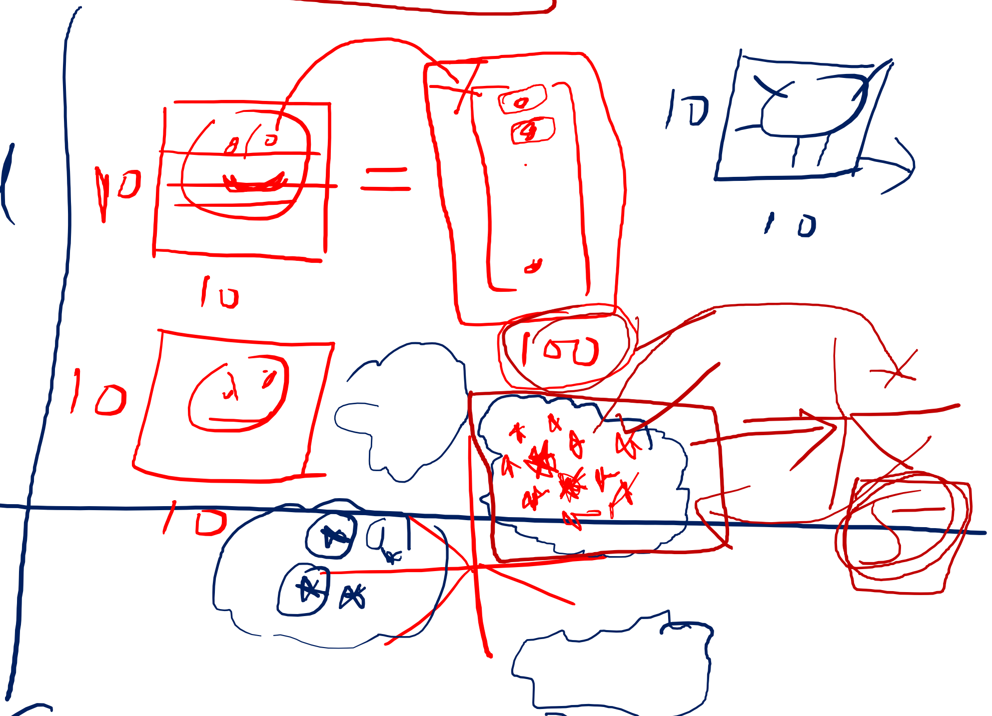
Probability of B occurring



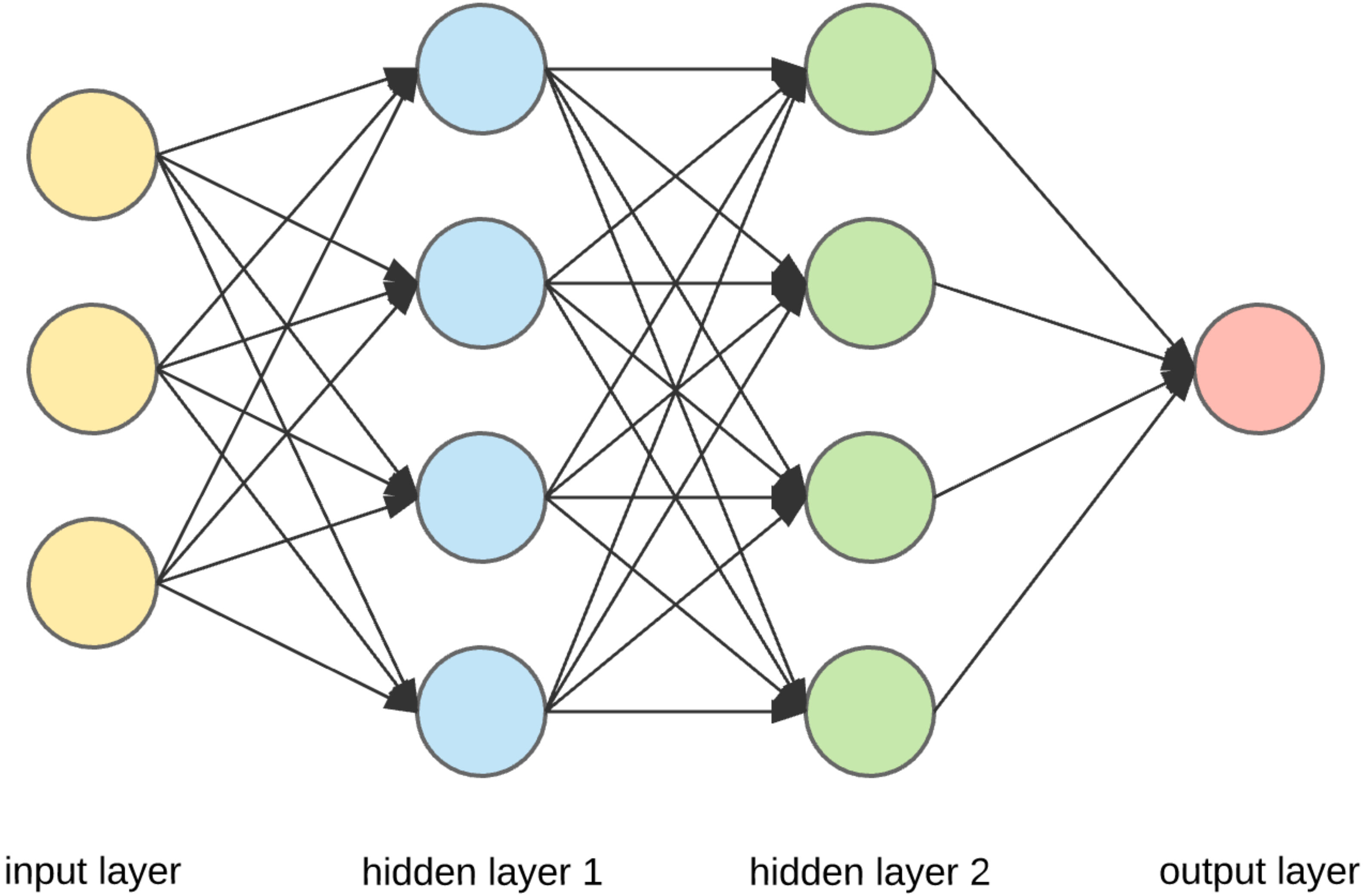
Dimensionality Reduction PCA

$\underset{N,d}{X} \rightsquigarrow \underset{N,k}{X_2}$, $k \ll d$

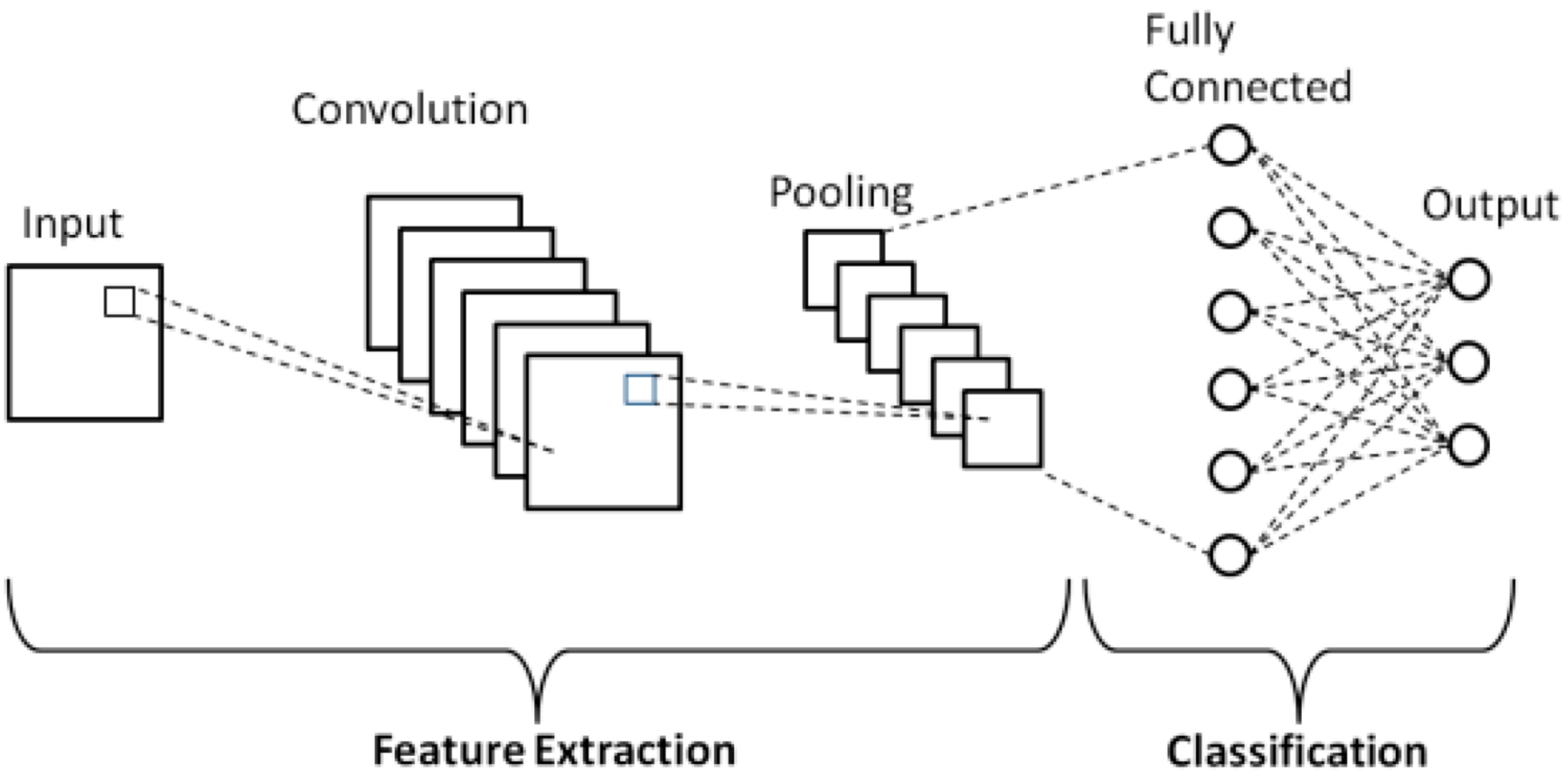
$\left(\underset{N,d}{X}, \underset{N,1}{y}\right) \xrightarrow{\text{FLD}} \left(\underset{N,k}{X_2}, \underset{N,1}{y}\right)$



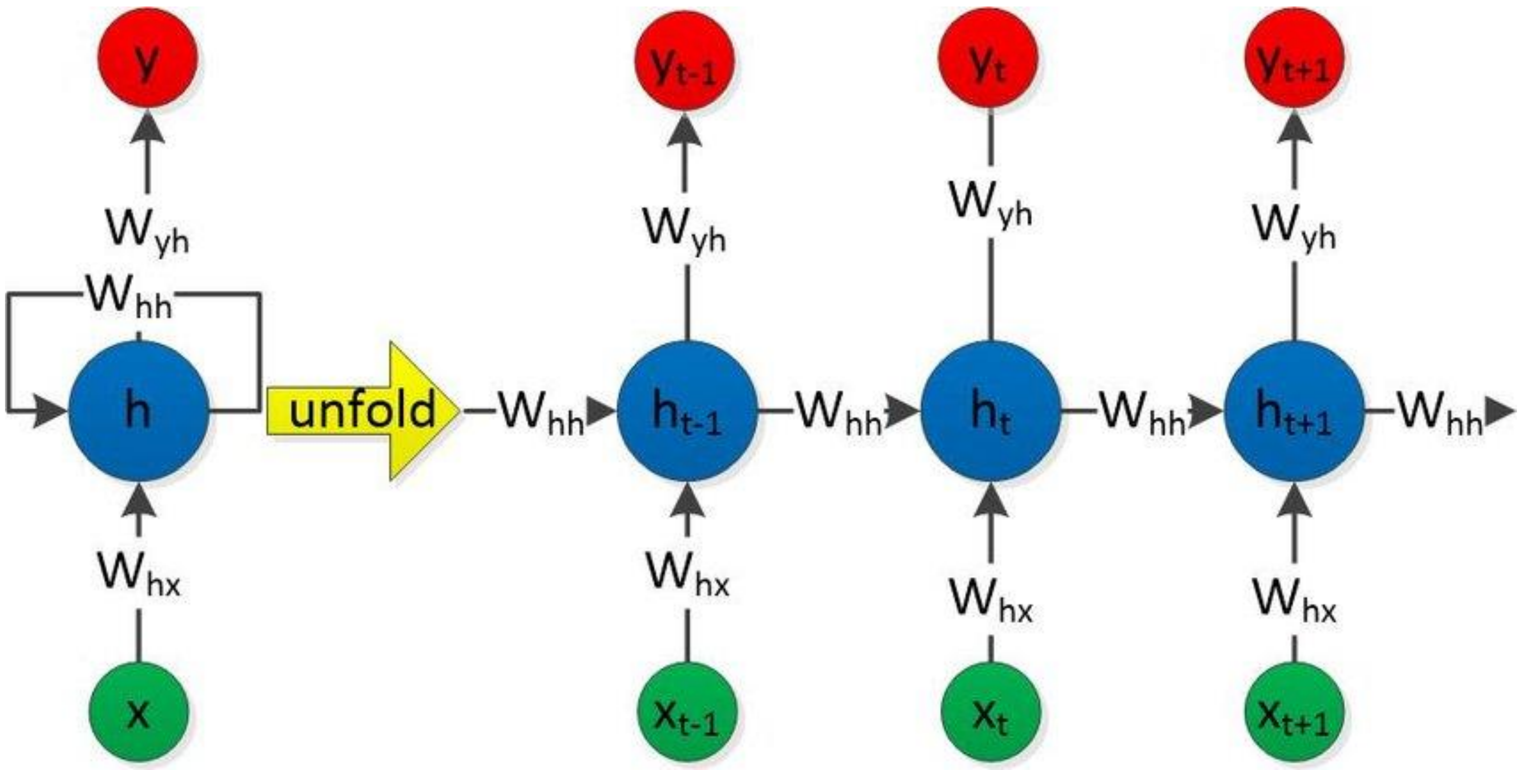
Deep Learning: ANNs



Deep Learning: CNNs



Deep Learning: RNNs



Mathematical wrap-up

$$f(x) = ax + b$$

$$w = \begin{bmatrix} a \\ b \end{bmatrix}$$

$$\cancel{f_w(x_i, y_i)}$$

$$\frac{1}{n}\sum_{i=1}^{n} L\left(f_w(x_i), y_i\right) + \lambda\, r(w)$$

Mathematical wrap-up

→ f, L, Preprocessing, regularizer

quantity λ:

$$Var = \sum_{i=1}^{N} (x_i - m_k)^2$$

Mathematical wrap-up

$$\sum_{i=1}^{N_k} (x_i - m_k)$$

$$V_k = \sum_{x \in C_k} (x - m_k)^{\boxed{2}}$$

$$\boxed{\sum_{i=1}^{K} V_i}$$

Mathematical wrap-up

$$f_w(x_i) = \underline{w_1} x_{i_1}^2 + \underline{w_2} x_{i_2}^2 + \underline{w_3} x_{i_1} x_{i_2} + w_4 x_{i_1} + w_5 x_{i_2} + w_6$$

$$\begin{bmatrix} x_{i_1}^2 & x_{i_2}^2 & x_{i_1} x_{i_2} & x_{i_1} & x_{i_2} & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \\ w_4 \\ w_5 \\ w_6 \end{bmatrix}$$

$$g_w(x_i) = w_1 x_{i_1} + w_2 x_{i_2} + w_3$$

$$\begin{bmatrix} x_{i_1} & x_{i_2} & 1 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix}$$

Mathematical wrap-up

$$\boxed{P(y_1=1 \mid x)} = \alpha_1$$

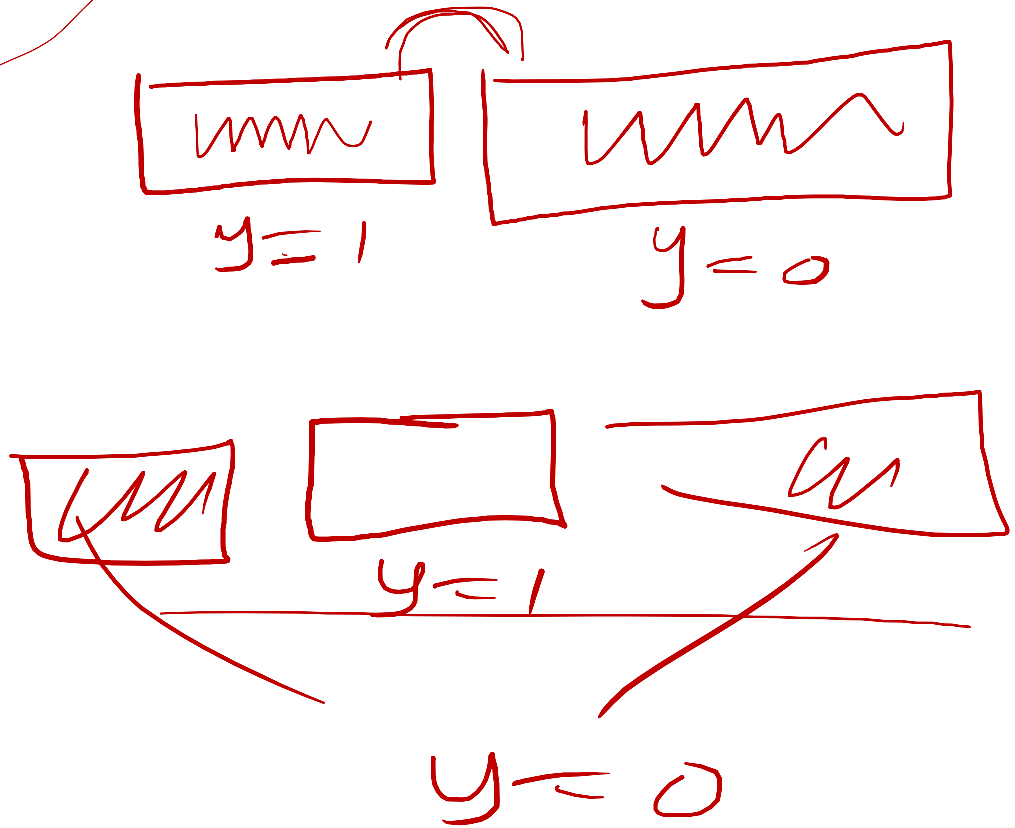
$$P(y=0 \mid x) = 1-\alpha$$

$$x_t$$

$$P(y_2=1 \mid x) = \alpha_2$$

$$P(y_3=1 \mid x)\ \alpha_3$$

$$=$$



Mathematical wrap-up

LDM :

$\cancel{w^T x}$

$$w = \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_d \end{bmatrix}$$

$$x_i = \begin{bmatrix} x_{i1} \\ x_{i2} \\ \vdots \\ x_{id} \end{bmatrix}$$

$$y_k = w_k^T x_i + w_{k0}$$

$$\sum_{i=1}^{N} \left(w_k^T x_i + w_{k0} - y_k \right)^2 + \lambda w_k^T w$$

Mathematical wrap-up

logistic regression

$$f_w(x_i) = \boxed{P_i} = P(y=1 \mid x_i)$$

$$P_i = \frac{1}{1+e^{-w^T x_i + w_0}}$$

→ Cross entropy

Mathematical wrap-up

$$P(y_i = 1 \mid x_i) = p_i$$

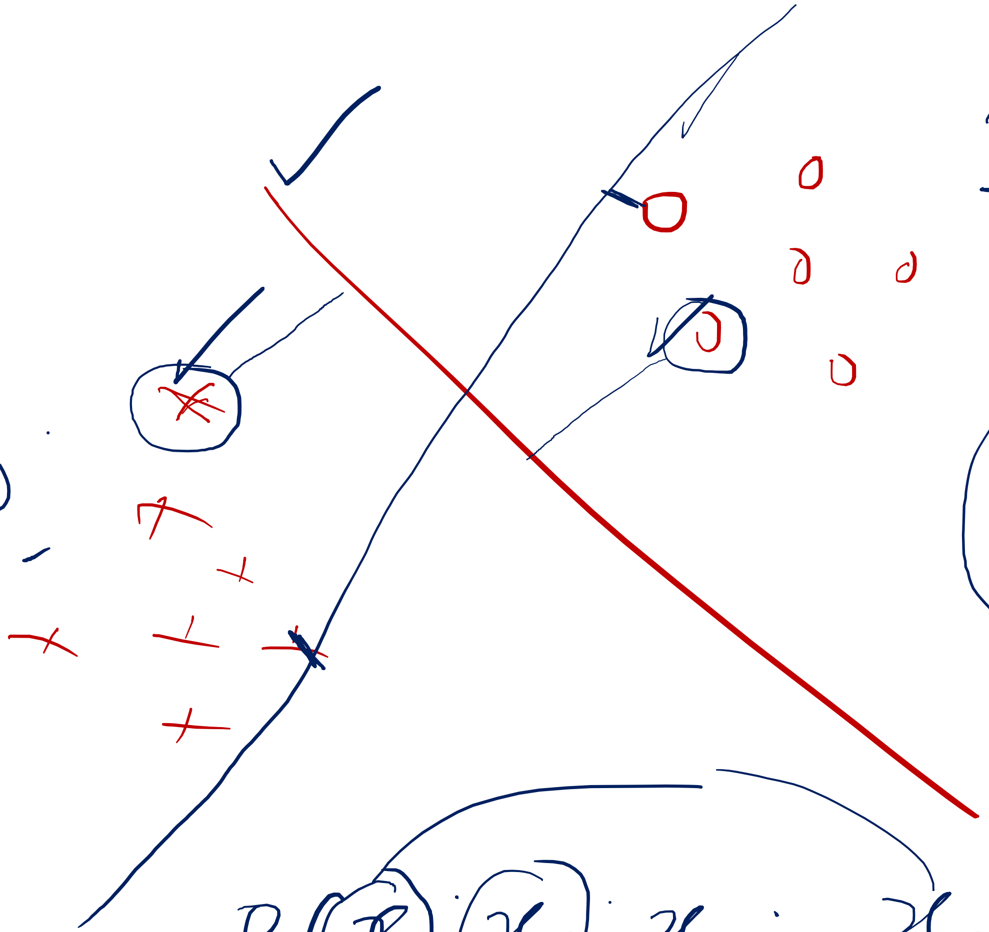
$$P(y_i = 0 \mid x_i) = 1 - p_i$$

$$p_i = \frac{1}{1 + e^{-w^T x_i + w_0}}$$

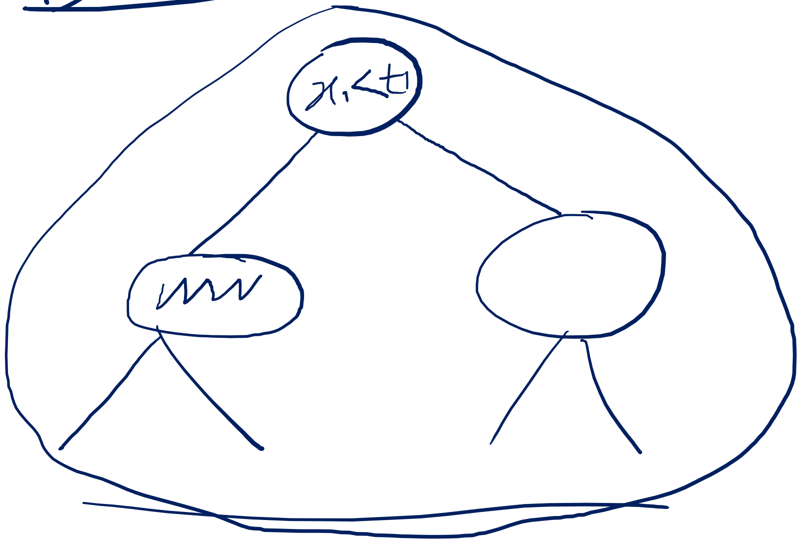
Mathematical wrap-up

SVM

Naive bayes



Decision Tree



$$P(x_{1i}, x_{2i}, x_{3i} \cdots x_{di} \mid y_i) = P(x_{1i} \mid y_i)\, P(x_{2i} \mid y_i) \cdots$$