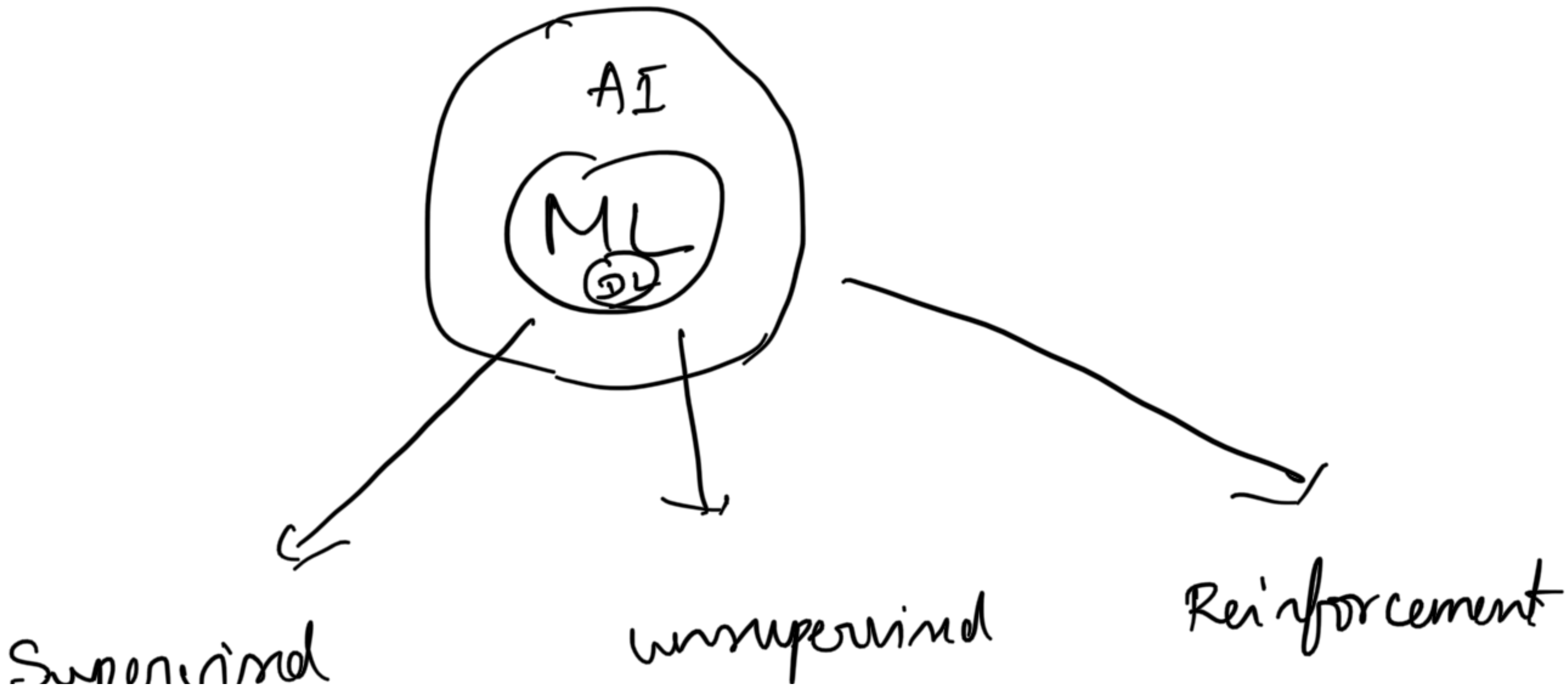
Outline

1. Overview of ML

2. Bayes classifier

3. Nearest neighbour classifier.



$\downarrow$

Data with Labels.

<u>Supervised Learning:</u> Data with labels.

$$(x_i, y_i)_{i=1}^{n} \longrightarrow n \text{ examples.}$$

Problem of Classifying some one as Child or Adult

$x_i = (\text{Height}_i, \text{weight}_i)$.

$y_i \in \{\text{Adult}, \text{child}\}$.

Features		Labels.
(5.1 ft, 60 Kg)	,	Adult
(5.0 ft, 50 Kg)	,	child

Data

$$(7.0\,\text{ft}, 60\,\text{Kg}) \xrightarrow{\text{Predict}} \text{child or Adult}$$

Supervised Learning

$\downarrow$

Classification

Task:

Features. Labels.

$$\left(x_i \in \mathbb{R}^d,\ y_i \in \{0,1\}\right)_{i=1}^{n}$$

Classifier

$$x \longrightarrow \boxed{\;h\;} \xrightarrow{\;y\;}$$

$y = h(x)$ Build h.

Random variable vs variable

Random variable is a variable.

because you have a distribution associated with values the variable can take.

$$X = \{ \underset{\downarrow \atop 0}{\text{Head}} , \underset{\downarrow \atop 1}{\text{Tail}} \}$$

$$\boxed{P(X=0) = \frac{1}{2} \qquad P(X=1) = \frac{1}{2}}$$

Distribution or frequency with values

1. Support of of a random variable
2. Distribution / Density associated with a random variable.

Examples of Random variables:

1. Throw of a die

$X \in \{1, 2, 3, 4, 5, 6\}.$ → Support of a RV

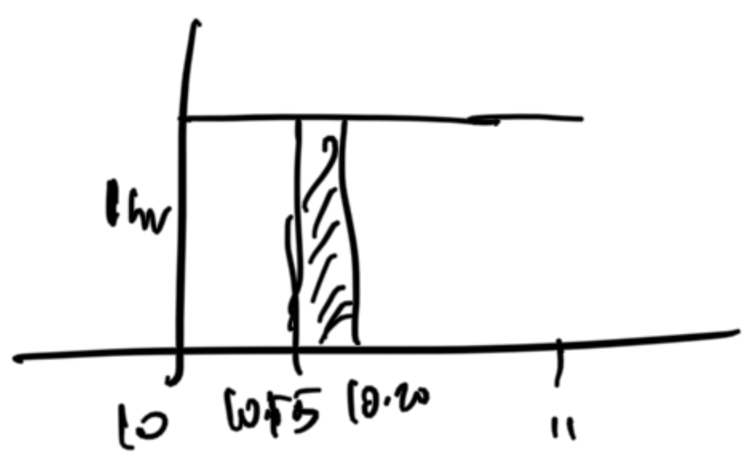
$P(X = i) = 1 \quad i = 1, 2, \ldots, 6.$

$P(X = i) = \frac{\quad}{6}$

Discrete Random variable.

2. Class session — (10 – 11.)

The arrival of a student to the class is a continuous random variable



mass density

m/length

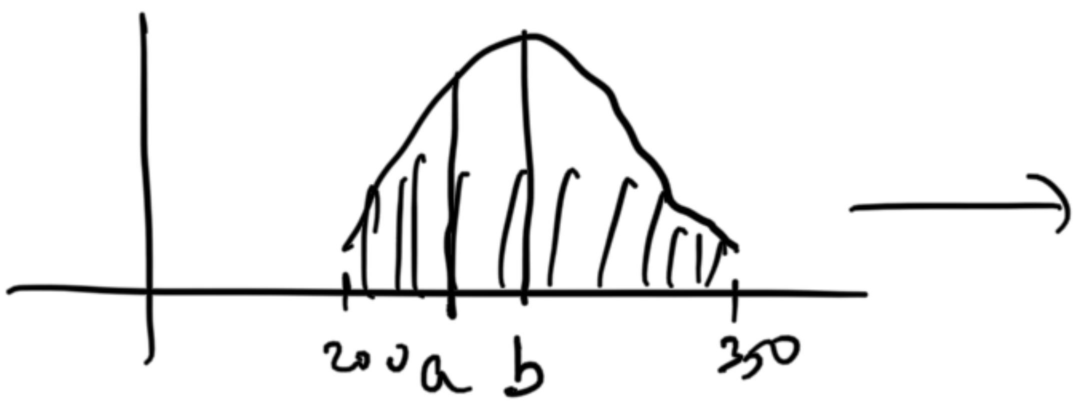
m/Area.

Continous random variable.

1. Time of arrival to class which is scheduled between 10 to 11 AM.

2. Temperature of a boiler.

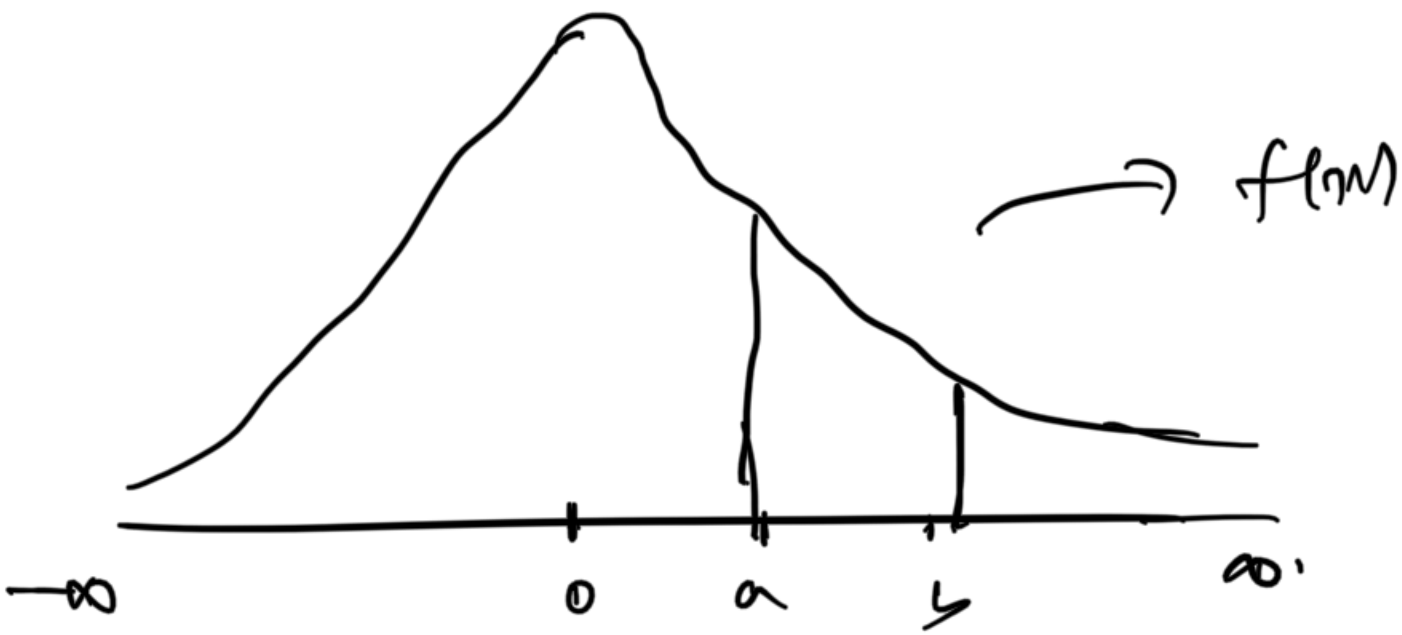
200 — 350



350

$$\int_{210} f(x)\, dx = 1.$$

$$P(a \leq x \leq b) = \int_a^b f(x)\, dx$$



$$\int_a^b f(x)\, dx = P[a \leq x \leq b].$$

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

$\downarrow$

Normal density.

Random variable (RV)

1. Discrete RV $\longrightarrow$ Probability mass function
2. Continuous RV $\longrightarrow$ Probability density function
3. Mixed RV

Bayes Rule: (Counting in two ways)

A, B → Events

$$\boxed{P(A|B)\,P(B) = P(B|A)\,P(A)}$$

Experiment	A	B
1. out 1	1	0
2. out 2	1	1
3. out 3	0	1
4. out 4	0	0
5. out 5	0	1

$P(A \cap B) = \frac{2}{7}$.

$P(A) = \frac{3}{7}$

$P(B) = \frac{4}{7}$

5\. Out ...

6\. Out 6 0 0

7\. Out 7 1 1

$$P(A|B) = \frac{2}{4}$$

$$P(B|A) = \frac{2}{3}$$

$$P(A)\,P(B/A) = \frac{3}{7} \times \frac{2}{3} = \frac{2}{7}$$

$$P(B)\,P(A/B) = \frac{4}{7} \times \frac{2}{4} = \frac{2}{7}$$

$$\boxed{\begin{aligned} P(A)\,P(B|A) &= P(B)\,P(A|B) \\ &= P(A \cap B) \end{aligned}}$$

Bayes Classifier.

$$P(D) = 0.2$$

$$P(R) = 0.8$$

Prediction strategies:

1. Predict he will default (always)
2. Predict he will Return (always).
3. with high P of coming

3. Coin with bias p ... heads.

$$p \longrightarrow \text{Return}$$

$$1-p \longrightarrow \text{Default.}$$

Strategy 1:

Predict always he will default

$$P(\text{error}) = 0.8$$

Strategy 2:

$$P(\text{error}) = 0.2$$

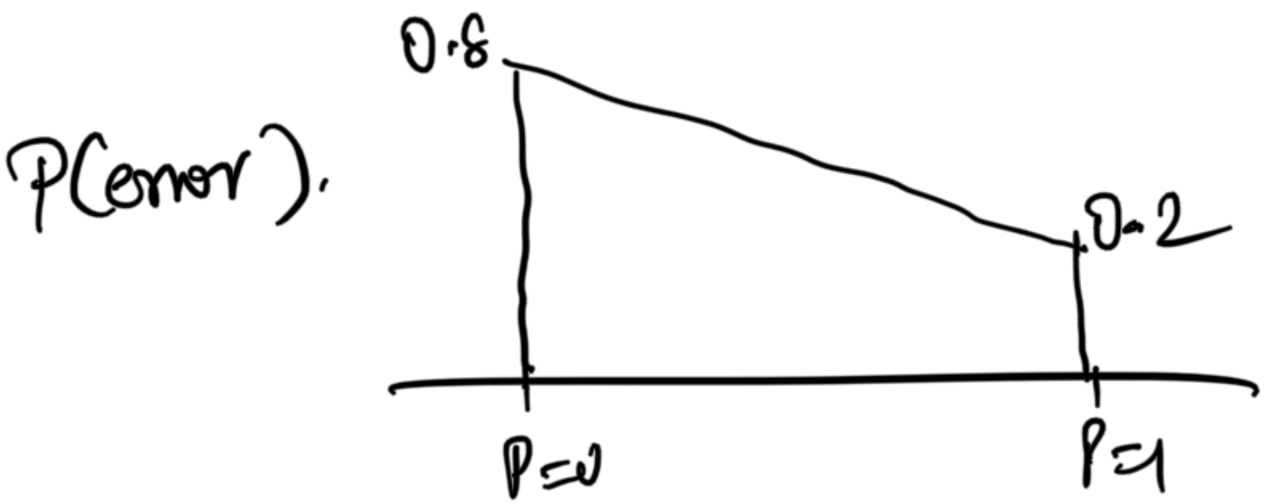
Strategy 3:

H $\quad p \longrightarrow$ Return

T $\quad 1-p \longrightarrow$ Default

$$P(\text{error}) = p\,0.2 + (1-p)\,0.8$$

$$= 0.8 - 0.6p$$



$P = 1 \Rightarrow$ Predict always Return.

H $\rightarrow$ more man 10 times his salary

L $\rightarrow$ less man 10 times his salary

$$\boxed{P(H \mid D) = \frac{2}{3}} \rightarrow \text{Given}$$

$$P(L \mid D) = \frac{1}{3}$$

$$\boxed{P(H \mid R) = \frac{1}{10}} \rightarrow \text{Given}$$

$$P(L \mid R) = \frac{9}{10}$$

$$P(D \mid H), \quad P(R \mid H)$$

$$P(D \mid L), \quad P(R \mid L).$$

Suppose,

$$P(D \mid H) > P(R \mid H)$$

$$H \longrightarrow D$$

$$P(D \mid H) < P(R \mid H)$$

$$P(D \mid L) < P(R \mid L)$$

$$L \longrightarrow R$$

H $P : D$	L $P : R$

$$P(D \mid H) = \frac{P(H \mid D)\,P(D)}{P(H)}$$

$$P(H) = P(H \cap D) + P(H \cap R).$$

$$= P(H|D)\,P(D) + P(H|R)\,P(R)$$

$$P(D|H) = 0.625$$

$$P(R|H) = 0.375$$

$$P(D|L) = 0.08$$

$$P(R|L) = 0.92$$

P(error) = P(High Loaner returns money)

+ P (low Loaner defaulting money)

$$= P(H)\,P(R|H) + P(L)\,P(D|L)$$

$$= P(R)\,P(H \mid R) + P(D)\,P(L \mid D)$$

$$\leq 0.2$$

<u>Bayes Classifier:</u>

C_0 $\rightarrow$ class 0 $\quad$ C_1 $\rightarrow$ class 1

$$\begin{cases} \text{Prior} & P(C_0),\ P(C_1) \\ \text{Likelihood} & f_0(x),\ f_1(x) \end{cases}$$

$$f_0(x) = P(X = x \mid C_0)$$

$$f_1(x) = P(X = x \mid C_1)$$

$f_1(x) = P(X = x \mid C_1)$

Posterior $\longrightarrow$ $P[C_0 \mid X=x]$

$$P[C_1 \mid X=x]$$

Bayes Classifier

$$P[C_0 \mid X=x] < P[C_1 \mid X=x]$$

$$x \longrightarrow C_1$$

$$P[C_0 \mid X=x] > P[C_1 \mid X=x]$$

$$x \longrightarrow C_0$$